CRM 91-221 / January 1992

Data Quality for the Helicopter Maintenance Phase of the Marine Corps Job Performance Measurement Project

Neil B. Carey
Catherine M. Hiatt

50 Years CNA 1992
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

Work conducted under contract N00014-91-C-0002.

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*

2. REPORT DATE
January 1992

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Data Quality for the Helicopter Maintenance Phase of the Marine Corps Job Performance Measurement Project

5. FUNDING NUMBERS
C - N00014-91-C-0002
PE - 65153M
PR - C0031

6. AUTHOR(S)
Neil B. Carey, Catherine M. Hiatt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Naval Analyses
4401 Ford Avenue
Alexandria, Virginia 22302-0268

8. PERFORMING ORGANIZATION REPORT NUMBER
CRM 91-221

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commanding General
Marine Corps Combat Development Command (WF 13F)
Studies and Analyses Branch
Quantico, Virginia 22134

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

All large-scale data collection efforts must contend with the issue of data quality. This research memorandum examines the quality of data collected for the helicopter maintenance portion of the Marine Corps Job Performance Measurement Project and describes measures taken to minimize the effort of questionable or missing cases.

14. SUBJECT TERMS
Accuracy, Aircraft maintenance, Correlation, Data acquisition, Helicopters, JPM (job performance measurement), Marine Corps personnel, Performance (human), Performance tests, Quality, Regression analysis, Statistical data, Test scores

15. NUMBER OF PAGES
78

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
CPR

18. SECURITY CLASSIFICATION OF THIS PAGE
CPR

19. SECURITY CLASSIFICATION OF ABSTRACT
CPR

20. LIMITATION OF ABSTRACT
SAR

50 Years CNA 1992 CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

21 February 1992

MEMORANDUM FOR DISTRIBUTION LIST

Subj: CNA Research Memorandum 91-221

Encl: (1) CNA Research Memorandum 91-221, *Data Quality for the Helicopter Maintenance Phase of the Marine Corps Job Performance Measurement Project*, by Neil B. Carey and Catherine M. Hiatt, Jan 1992

1. Enclosure (1) is forwarded as a matter of possible interest.

2. All large-scale data collection efforts must contend with the issue of data quality. This research memorandum examines the quality of data collected for the helicopter maintenance portion of the Marine Corps Job Performance Measurement Project and describes measures taken to minimize the effect of questionable or missing cases.

Lewis R. Cabe
Director
Manpower and Training Program

Distribution List:
Reverse page

Subj: Center for Naval Analyses Research Memorandum 91-221

Distribution List

SNDL

45A2 CG I MEF
45A2 CG II MEF
45A2 CG III MEF
46B CG FIRST MAW
46B CG SECOND MAW
46B CG THIRD MAW
A1 DASN - MANPOWER (2 copies)
A1H ASSTSECNAV MRA
A2A CNR
A4 PERS-11
A5 PERS-2
A5 PERS-5
A6 CG MCRDAC - WASHINGTON
A6 HQMC AVN
A6 HQMC MPR & RA
Attn: Code M
Attn: Code MR
Attn: Code MP
Attn: Code MM
Attn: Code MA (3 copies)
Attn: Code MPP-54
FF38 USNA
Attn: Nimitz Library
FF42 NAVPGSCOL
FF44 NAVWARCOL
FJA1 COMNAVMILPERSCOM
FJA13 NAVPERSRANDCEN
Attn: Technical Director (Code 01)
Attn: Technical Library
Attn: Dir, Manpower Systems (Code 11)
Attn: Dir, Personnel Systems (Code 12)
Attn: Dir, Testing Systems (Code 13)
Attn: CAT/ASVAB PMO
FJB1 COMNAVCRUITCOM
FT1 CNET
V12 CG MAGTEC
V12 CG MCCDC
Attn: Studies and Analyses Branch
Attn: Director, Warfighting Center
Attn: Warfighting Center, MAGTF Proponency and Requirements Branch (2 copies)
V12 CG MCRDAC - QUANTICO

OPNAV

OP-01

CRM 91-221 / January 1992

Data Quality for the Helicopter Maintenance Phase of the Marine Corps Job Performance Measurement Project

Neil B. Carey
Catherine M. Hiatt

Operations and Support Division

50 Years CNA 1992
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

All large-scale data collection efforts must contend with the issue of data quality. This research memorandum examines the quality of data collected for the helicopter maintenance portion of the Marine Corps Job Performance Measurement Project and describes measures taken to minimize the effect of questionable or missing cases.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The Marine Corps Job Performance Measurement (JPM) Project is a large-scale effort to validate the Armed Services Vocational Aptitude Battery (ASVAB) against measures of job performance. For the helicopter maintenance phase, over 650 helicopter mechanics—in Military Occupational Specialties (MOSs) 6112, 6113, 6114, and 6115—were tested for two days each on a variety of performance measures. In addition to taking an eight-hour hands-on test of mechanical performance, examinees took a paper-and-pencil job knowledge test (JKT), performed certain portions of the General Aptitude Test Battery (GATB), took the ASVAB by computer (CAT-ASVAB), and were administered several new computerized predictors of job performance (ECAT).

The volume of data was enormous. Many problems could potentially affect the quality or completeness of data so precautions were taken to minimize the possibility of poor or missing data. Test administrators were trained extensively, data were checked for completeness every day, equipment was checked every day, and the consistency of responses was monitored daily. Examinees were briefed on the importance of giving their full effort. Despite these precautions, there were still individual cases in which the accuracy of the data was questionable and other instances in which the data were incomplete.

This memorandum quantifies the amount of questionable or incomplete data and details the procedures used to minimize the effect such data would have on later analyses.

IDENTIFICATION OF UNUSUALLY LOW SCORES

Occasionally, a test may fail to measure properly the ability of a particular person even though the test may provide excellent measurement for a group. For such persons, it is possible that some condition occurred that produced unusually low scores (e.g., lack of motivation, illness, lack of sleep, inattentive marking of the answer sheet, random responses, or application of the wrong answer key).

To identify unusually low scores, one must compare scores with those achieved when the person was tested under motivated conditions. Enlistment ASVAB scores reflect performance when highly motivated because enlistment scores determine eligibility for the Marine Corps. Persons whose job knowledge test scores were far below what would be expected on the basis of their enlistment ASVAB scores were assumed to have given less than their full effort on the JKT.

Decision rules were established for the identification of unusually low scores based on the prediction of job knowledge scores from enlistment ASVAB and time in service. Given these criteria, six scores were declared aberrant for the JKT across the four MOSs. Deleting these aberrant scores increased the means for the test and decreased the standard deviation. The correlation of the JKT with the enlistment Mechanical Maintenance (MM) aptitude composite scores increased slightly. These changes in sample statistics indicated that the deleted scores were typically outlier cases.

To check for Marines whose CAT-MM scores were not reflective of their full effort, CAT-ASVAB MM scores were predicted from enlistment MM scores. As with the JKT, persons whose CAT-ASVAB scores were far below what would be expected on the basis of their enlistment scores were assumed to have given less than their full effort on the CAT-ASVAB. In such cases, the CAT-ASVAB score is not reflective of the person's true aptitude.

The CAT-ASVAB scores of Marines with extremely low CAT-MM scores relative to their enlistment MM were deleted. Across the four MOSs, five scores were deleted for this reason. Deleting these scores increased the mean and decreased the standard deviation. Correlations with enlistment MM increased slightly, indicating that deleted cases were outliers.

To determine whether some Marines did not give their full effort in taking the hands-on performance test (HOPT), HOPT was predicted from enlistment MM and time in service. Three HOPT scores from MOS 6113 were deleted because they were extreme outliers in this analysis. No HOPT scores were deleted for any of the other three MOSs.

IMPUTATION OF MISSING DATA

Hands-on performance data were collected at the step level; a person either passed or failed to perform a specific action. Steps were aggregated to form task scores. It was not always possible to collect complete information for each person—there were over 200 steps for each MOS's hands-on test. Examinees could have incomplete data as a result of weather conditions, equipment failure or unavailability, being called away before completion, or performing a step that was unobservable to the test administrator.

Despite the many ways data could be missing, very few data were missing. Complete data were collected for 95 percent of all tasks administered.

Imputation is the process of estimating the score that would have occurred if circumstances had not prevented actual scoring. Imputation was performed to make fullest use of the data. Data were imputed at the step level. Sample statistics for all variables with complete information before the step-level imputation were compared to the sample statistics after imputation. The shifts in mean performance scores were relatively small compared to the standard deviation of performance scores. With few exceptions, correlations of performance with aptitude changed insignificantly as a result of imputation.

PROBLEM LOGS

Problem logs, maintained by field data collectors, recorded instances of difficulty in collecting or maintaining quality of data. Problem logs were kept for the job knowledge test, CAT-ASVAB, ECAT, GATB, and two "administrative duties" tests. Field data collectors' comments noted that data were lost or questionable because of lack of effort from examinees or situational disruptions. Considering that data were collected over a period of several months in two different locations, the logs indicated relatively few problems.

There were very few missing cases for any of the data sources covered by the problem logs. Inspection of the actual data confirmed the small number of problem log entries: Across the four MOSs, the actual amount of fully missing data ranged from a high of 24 cases for the CAT-ASVAB (3.6 percent) and 21 cases for ECAT (3.2 percent) to a low of no missing data for the JKT.

Further analyses focused on problems with the CAT-ASVAB and ECAT data collection. Most reported problems for the CAT-ASVAB were caused by power outages and problems with the power cord during test administration. It was noteworthy, however, that the largest problem of missing data—overflow of data on ECAT disks at site B—was not mentioned in the problem logs; the test administrators did not discover the problem until later. Test administrators' lack of awareness of this problem suggests that better computer training is needed for future ECAT data collection. Other problems with the ECAT involved broken response pedestals. This finding suggests that substantial numbers of replacement pedestals must be kept for administration of the ECAT.

RESULTS

Relatively few unusually low scores were observed for the JKT. The aberrant data cases were outliers, so their deletion generally improved sample correlations and reduced standard deviations. The criteria for identifying unusually low scores were specifically chosen to be conservative. Specific deletions were confirmed against other information whenever possible. Less than 1 percent of JKT scores were deleted as a result of these procedures. Given the verification across different information sources (residual analysis, percent correct score, problem logs, personal biserial correlation), few people, if any, should have been misidentified as having aberrant scores when, in fact, the test score was a reasonable estimate of their ability.

Across the four MOSs, 74 tasks composed of 1,014 steps were administered to 658 Marines. Complete data were collected for 95 percent of all tasks. All cases were deemed recoverable by imputation of missing data. Sample statistics were insignificantly affected by imputation. Indeed, this was the intended outcome sought by employing an imputation procedure that incorporated steps to minimize the impact of imputed values.

As a result of these data quality analyses that identified unusual response patterns and imputed missing data for the helicopter maintenance JPM data, further analytic investigations can proceed with confidence in the soundness of the data and the integrity of the results.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

THIS PAGE INTENTIONALLY LEFT BLANK

TABLES

TABLES (Continued)

INTRODUCTION

The Job Performance Measurement (JPM) Project is a long-term effort to validate the Armed Services Vocational Aptitude Battery (ASVAB) against hands-on measures of job performance. The Helicopter Mechanical Maintenance phase of the project tested over 650 mechanics in four military occupational specialties (MOSs) at two test sites. The mechanics' job duties included repair of the CH-46 (MOS 6112), CH-53A/D (MOS 6113), U/AH-1 (MOS 6114), or CH-53E (MOS 6115). Each mechanic was required to complete 15 or more hands-on mechanical tasks on the MOS-specific helicopter.

Properly implemented hands-on tests are very resource intensive: Test administrators must individually observe and score performance of job tasks, so administration costs are considerably higher than for paper-and-pencil testing. Test administrators must have extensive training and be given frequent feedback on their performance judgments. Helicopters must be restored to proper condition before the next mechanic performs a task. Because hands-on tests require a well-organized flow of examinees, hands-on testing requires considerable attention to the assignment and transportation of personnel.

The JPM project also addressed other manpower research issues, such as whether less expensive "surrogate" measures could be used in place of hands-on tests and whether "new predictors" could enhance the predictive power of the ASVAB. Therefore, examinees were administered a paper-and-pencil test of helicopter mechanic job knowledge, were administered the ASVAB by computer (CAT-ASVAB), and were given a series of computerized new predictors designed to enhance the ability of the ASVAB to predict performance. Portions of the General Aptitude Test Battery (GATB) that measure manual dexterity were also administered as part of this project.

The volume of data was enormous, as was the potential for problems to affect the quality of data or their completeness. Test equipment could break, test sites' equipment setups could differ, a test administrator could fail to see whether a step was completed, or an examinee could fall ill or be called away. Even if these problems did not occur, other situations could affect the quality of data. Examinees could rush through the test, failing to give their full effort; computers, disks, or information networks could fail; the wrong test form could be given; or unavoidable distractions could upset testing.

Because of the many problems that may beset large-scale data collection efforts, significant precautions were taken to minimize the possibility of poor and missing data [1]. Preliminary tryout testing and a command review were conducted for all tasks on the performance test. Data quality was affected by the extent to which Marines were motivated to attempt all performance measures. To ensure full effort, each participant was given an illustrated pamphlet that explained the importance of the study to the future of the Marine Corps. Immediately before testing began, a Marine Corps officer gave a short talk that emphasized the importance of giving one's full effort.

Continuous monitoring during the testing identified potential problems so they could be corrected as soon as possible. Data quality was monitored daily through verification[1] of answer sheets, daily entry of all hands-on responses, and maintenance of problem logs to identify specific problem cases. To encourage effort in taking the CAT-ASVAB, Marines were informed that their scores of record would be changed to their CAT-ASVAB scores if their CAT-ASVAB scores exceeded their scores of record. This could have significant payoff for people who wanted to transfer to other occupational fields with higher aptitude requirements. In research on the earlier infantry phase of this study, this incentive appeared to be effective [2]. No changes would be made if the CAT-ASVAB scores were lower than the scores of record.

Precautions were also taken to minimize the amount of missing data. A Marine Corps technician was available each day so that mechanical problems and part failures could be dealt with promptly. Test administrators were instructed in ways to observe all steps being performed. Data were reviewed daily for completeness, and examinees were scheduled to retake any portions of the test they missed.

Despite these initial tryouts and quality-control procedures, there were individual cases in which the accuracy of the data was questionable and other cases in which the data were simply incomplete. Both of these factors affect overall data quality and can affect analyses yet to be conducted on JPM data. For example, Maier [3] has found that data quality-control procedures can make large differences in the computed validity coefficients.

To identify data inaccuracies at the individual level, item responses were examined for unusually low scores compared to their scores earned under motivated

1. The onsite manager verified answer sheets by reading each sheet. If there were inconsistent or missing data, the manager questioned the test administrator and made whatever corrections were necessary.

conditions. Enlistment ASVAB scores reflect motivated performance because such scores determine one's eligibility for the Marine Corps. An unusually low score compared to one's enlistment Mechanical Maintenance (MM) score could indicate fatigue, illness, random guessing, or unknowingly skipping a question so that responses were always meant for the adjacent item. Such causes are unrelated to a person's ability, so they are errors for purposes of this project. Data that have such unusual responses must be declared missing.

An examinee might have incomplete data for a number of reasons. The Marine might be called away for an emergency, fall ill, or experience equipment breakage, or the test administrator might be unable to observe the examinee's response. Such events and conditions are not under the control of the examinee and, hence, are considered random. Data that are missing because of such random events can be estimated using data from other parts of the test. For examinees with partially missing data, some data are better than no data, and the missing data can be estimated using available data. Specific procedures were developed for the estimation of missing data at the step level.

Given that JPM analyses, yet to be conducted, are sensitive to outliers and require complete information, this research memorandum presents the procedures used to ensure the quality and completeness of the MM data. Methods for identifying unusual response patterns are described, and deletion of individual cases is justified based on checks between different sources of information. The magnitude of missing data at the step level is presented. The impact of deletion of aberrant data and imputation of missing data are documented by noting changes in the sample descriptive statistics.

As part of the data collection procedures, test administrators kept daily problem logs that noted anomalies in testing procedures. Results from the daily problem logs are analyzed to determine common problems in the collection of each type of data (e.g., CAT-ASVAB, new predictors, and job knowledge test) and to make recommendations for future data collection efforts.

IDENTIFICATION OF UNUSUALLY LOW SCORES

A test can fail to measure one person's ability even though the test adequately measures abilities in the group. For example, a person's scores might be anomalous because of lack of effort, cheating, random guessing, or unknowingly skipping a question. Such occurrences guarantee that the test is not properly measuring the person's abilities.

CAT-ASVAB

This project obtained two sets of ASVAB scores. **Enlistment scores** were obtained under motivated conditions, when the person's score affected acceptance into the Marine Corps, so these scores are assumed to be accurate reflections of aptitudes. **CAT-ASVAB scores**, obtained during JPM testing, could indicate less motivation because the examinee was already accepted into the Marines.

To detect examinees whose CAT-MM scores did not reflect their actual abilities, CAT-ASVAB MM scores were predicted from enlistment MM scores[1] using linear regression. Discrepancies between the actual CAT-ASVAB score and the score predicted from the enlistment score are called residuals. Further details concerning linear regression and residuals can be found in appendix A. Large negative residuals identified people whose CAT-ASVAB scores were not accurate indicators of their aptitude. Residuals were computed from the regression and plotted as shown in figure 1. CAT-ASVAB scores more than three standard residuals below what would be predicted from enlistment scores were considered indicative of poor motivation and were dropped from further analyses. Using this criterion, three CAT-MM scores were dropped for MOS 6112, one was dropped for MOS 6113, and one was dropped for MOS 6115. No CAT-MM scores were deleted for MOS 6114.

---

1. More accurate predictions could be made by using all ten subtests of the ASVAB, rather than just the MM composite, which uses only four subtests. Therefore, results using all ten subtests were compared with those using just enlistment MM. Using the ten subtests, the adjusted R^2 increased from .46 to .58 for MOS 6112, from .30 to .46 for MOS 6113, from .41 to .44 for MOS 6114, and from .33 to .53 for MOS 6115. However, the two methods yielded nearly identical results: Across the four MOSs, only two outliers (one in MOS 6112 and one in MOS 6113) identified by enlistment MM were not identified as an outlier when all ten subtests were used, and those scores barely missed the standardized residual cutoff of −3.0. One score identified as an outlier using all ten subtests—in MOS 6114—was not identified using enlistment MM alone. That score barely missed the cutoff using enlistment MM. It was decided to use enlistment MM for all outlier analyses because 85 examinees lacked subtest scores but had enlistment MM scores. Including these 85 examinees in the analysis resulted in one outlier (in MOS 6112) being identified who otherwise would not have been discovered because of lack of all ten subtest scores.

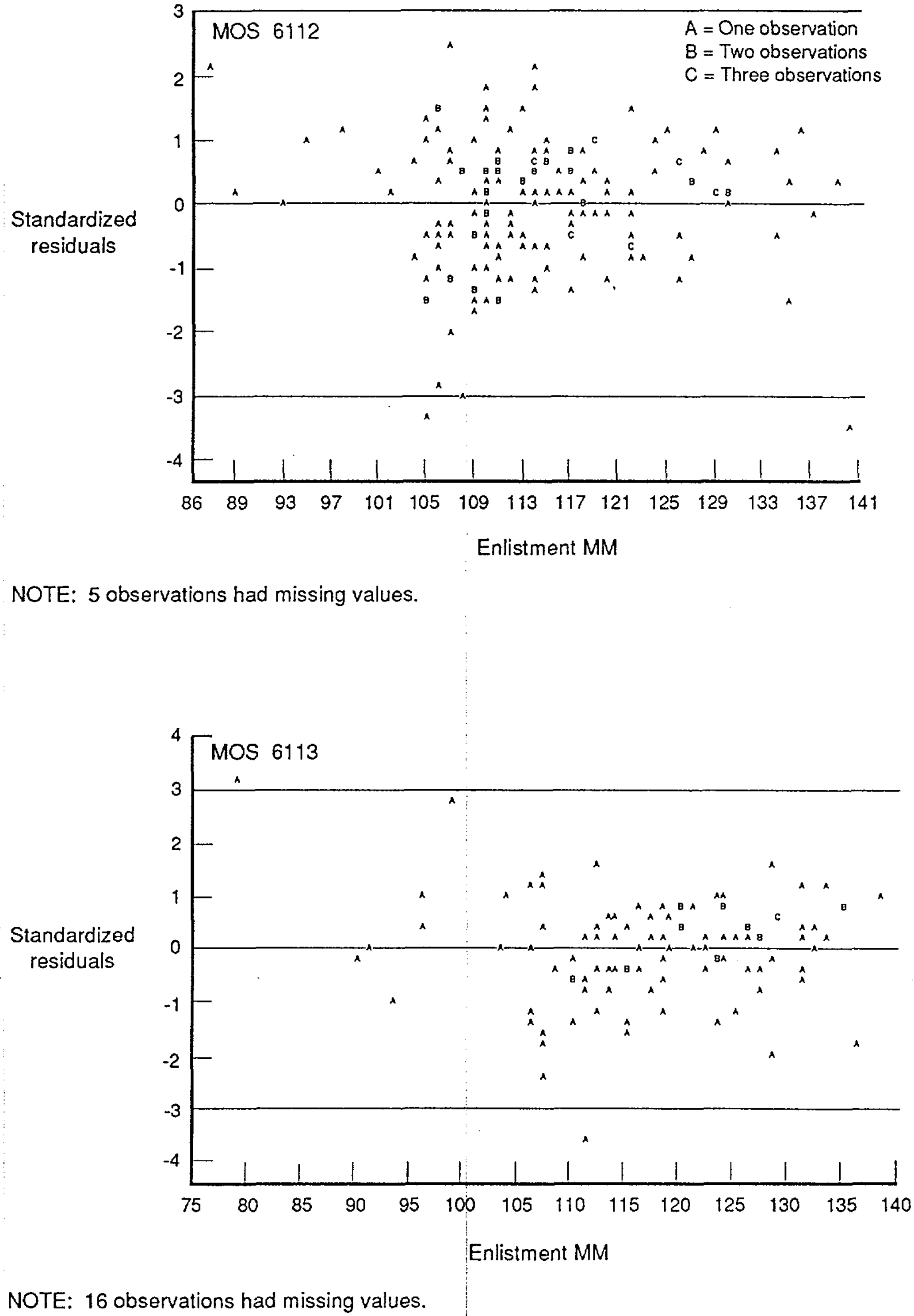
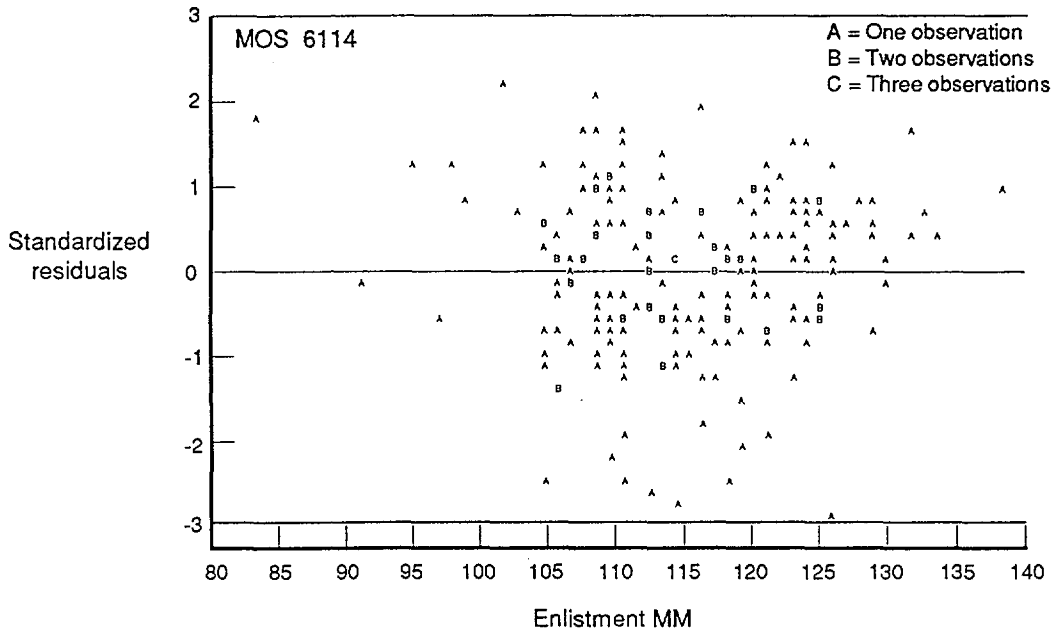
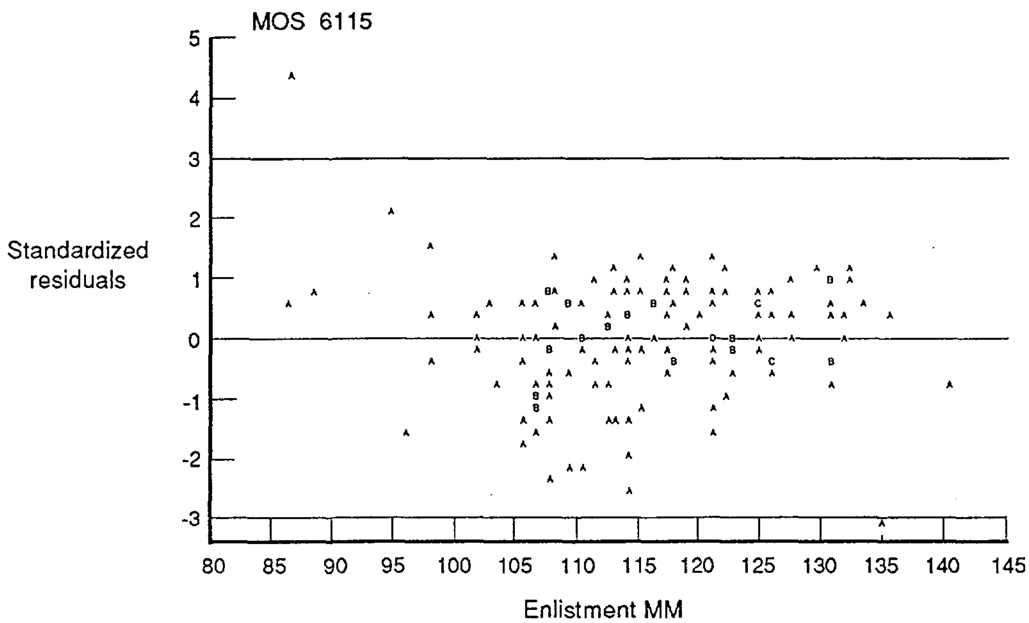


Figure 1. Residuals from regression of CAT-ASVAB MM composite on enlistment MM



NOTE: 16 observations had missing values.



NOTE: 9 observations had missing values.

Figure 1. (Continued)

Job Knowledge Test (JKT)

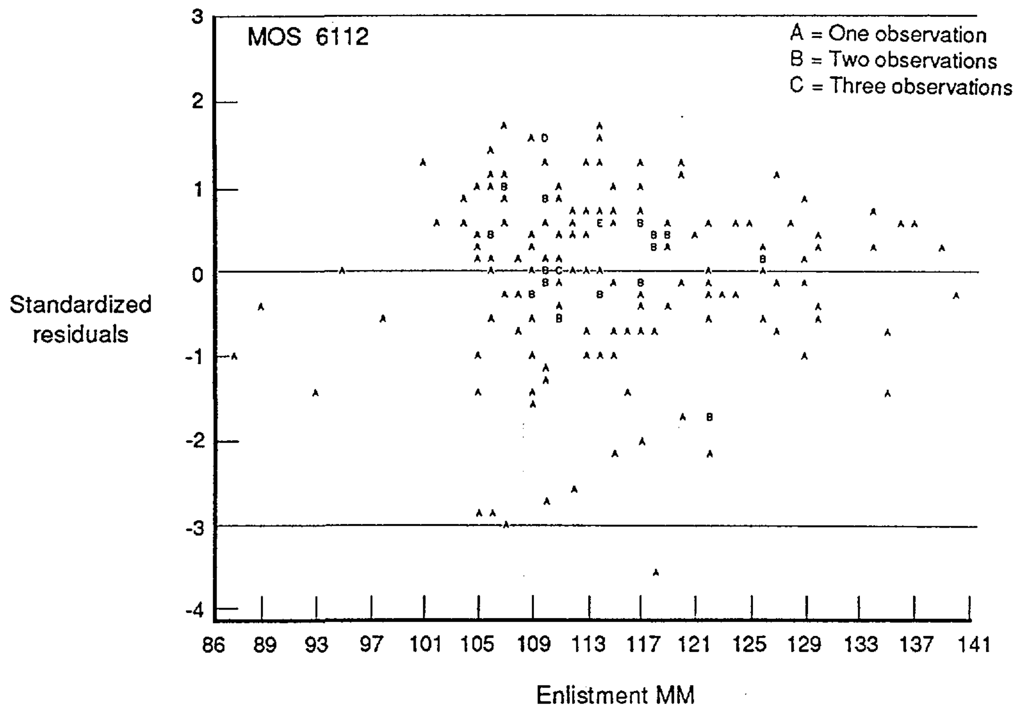
Three quality checks were conducted for the JKT: verification of test form, analysis of item quality, and identification of people with questionable job knowledge scores.

Two forms of the JKT were administered. To verify the form code for each written test (or to determine a form code if one was not marked), all answer sheets were scored against both answer keys. To verify the correct form code, comparisons were made of individual total scores resulting from each answer key. A higher total score indicated the correct test form. For borderline cases where the total score was the same for both forms, the reported form was used. Using this method, 14 examinees' scores were changed: seven for MOS 6112, none for MOS 6113, six for MOS 6114, and one for MOS 6115. Gains resulting from changing the form code ranged from 4 points to 61 points; the average gain was over 38 points.

To assess the measurement quality of items on the JKT, item point biserial correlations with total test score were computed. Point biserial correlation is the relationship between the scored item response (correct or incorrect) and the total test score. Positive values indicate that the item is probably functioning properly; negative correlations indicate possibly miskeyed items or poorly worded item alternatives. These analyses identified 12 items with negative correlations that had apparently been miskeyed (1 of the 179 items for MOS 6112, 2 of the 161 items for MOS 6113, 5 of the 168 items for MOS 6114, and 4 of the 165 items for MOS 6115). Once the key errors were corrected, 15 items still appeared questionable because of very low correlations with total score or near-chance levels of correct responses (three items for MOS 6112, six items for MOS 6113, three items for MOS 6114, and three items for MOS 6115). Upon consulting with subject matter experts, it was discovered that these items were ambiguous, so they were dropped from further analyses.

To identify people with questionable JKT results, scores were compared with those taken by the examinees under motivated conditions. To identify those who gave less than full effort to the JKT, scores were predicted from enlistment scores and time in service[1] using linear regression. Large negative residuals identified people whose job knowledge scores were not accurate indicators of their knowledge. Residuals were computed from the regression and plotted as shown in figure 2. One

1. Time in service (TIS) was used as a predictor for JKT because job knowledge typically increases with experience. TIS was not used as a predictor for CAT-ASVAB because experience is generally not related to mental aptitude.



NOTE: 4 observations had missing values.

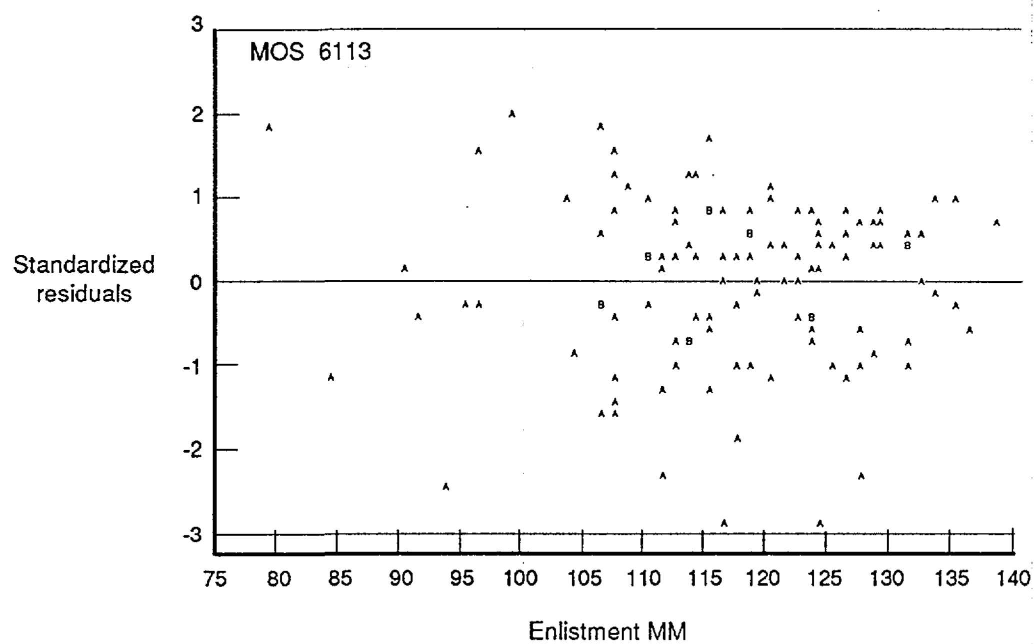
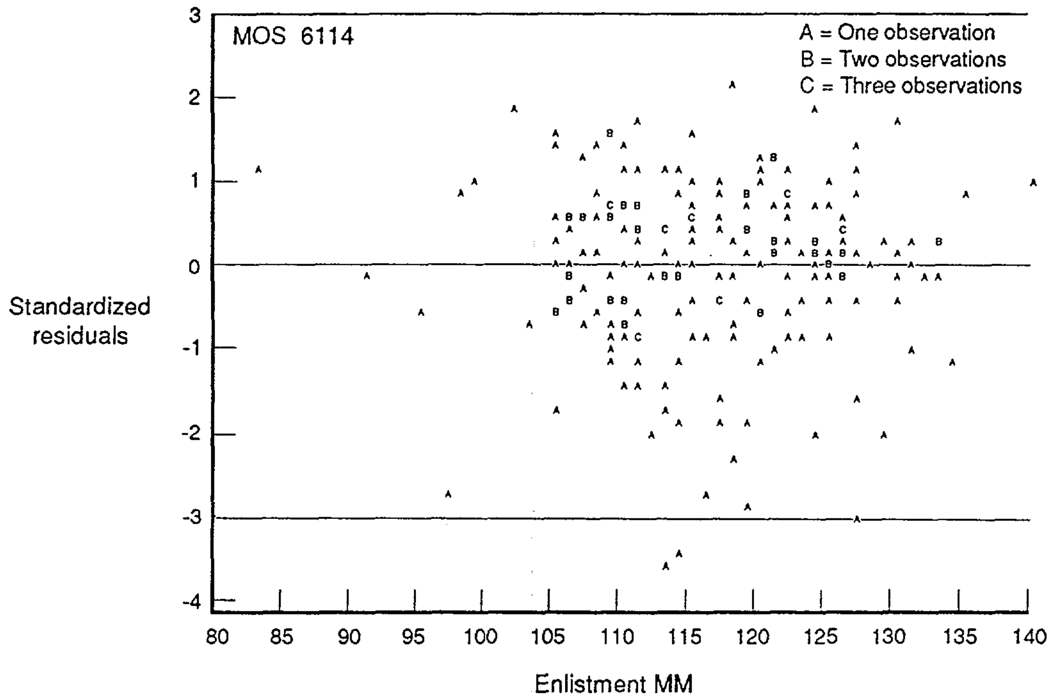


Figure 2. Residuals from regression of JKT on enlistment MM and time in service



NOTE: 3 observations had missing values.

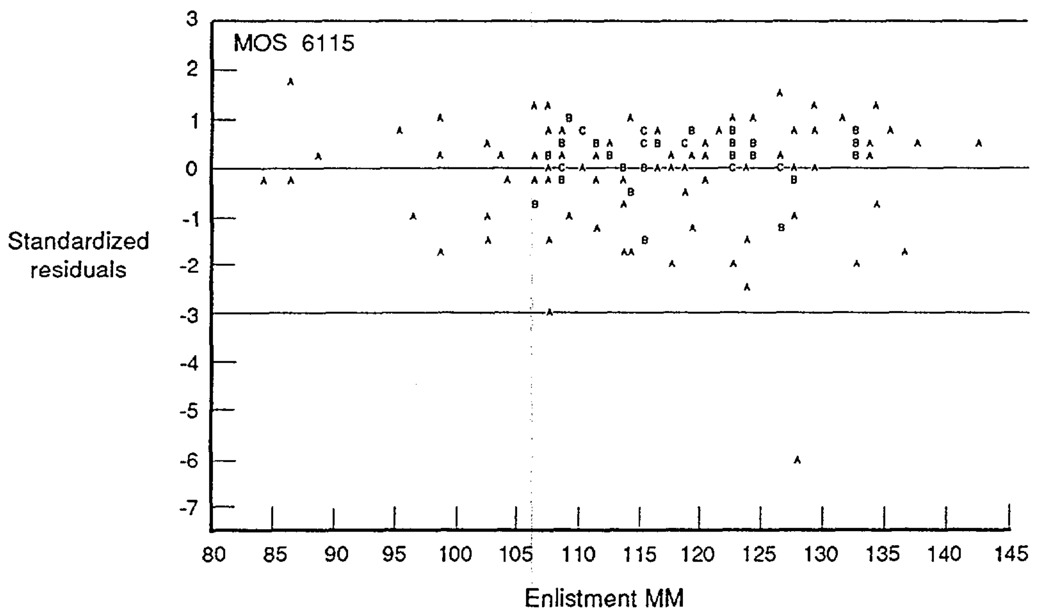


Figure 2. (Continued)

decision rule was established for the identification of unusually low JKT scores: standardized residual < = –3.0. This critical region for defining aberrant scores is noted on the figure.

Based on these criteria, six scores were declared aberrant for the JKT (two for MOS 6112, three for MOS 6114, and one for MOS 6115). As a result of deleting the six aberrant scores, the mean for the JKT increased and the standard deviation went down. The correlation of this test with the Mechanical Maintenance (MM) aptitude composite score increased slightly for all MOSs for which scores were deleted (table 1). These changes in sample statistics indicate that the scores that were deleted were typically aberrant cases.

Table 1. Change in sample statistics caused by deleting aberrant scores of JKT

					Correlations with	
	MOS	n	Mean	Std. dev.	Hands-on total score	Enlistment MM score
6112						
	Before deletion	174	69.8	13.1	.58	.30
	After deletion	172	70.3	12.4	.61	.31
6113						
	No deletions					
6114						
	Before deletion	215	72.6	9.6	.50	.24
	After deletion	212	73.0	9.1	.51	.26
6115						
	Before deletion	149	79.4	9.9	.46	.08
	After deletion	148	79.9	8.5	.52	.14

Hands-On Performance Test (HOPT)

Enlistment MM scores reflect motivated test taking because enlistment scores determine eligibility for the Marine Corps. Although every attempt was made to ensure full effort for JPM testing, some examinees might not have been motivated to perform. Therefore, HOPT scores were regressed on enlistment MM[1] and time in service to determine whether some HOPT performers might have been unmotivated. HOPT scores far below those predicted by enlistment scores and time in service would be candidates for deletion. Three HOPT cases were deleted after examination of residuals and inspection of the problem logs, all in MOS 6113. In all cases, the deleted examinees' scores had residuals less than –2.5.

1. Using all ten ASVAB subtests instead of just enlistment MM did not change which HOPT performers were identified as unmotivated.

IMPUTATION OF MISSING DATA

Data collected for the Marine Corps JPM Project were extremely difficult and expensive to obtain. Despite the best of intentions, it was not always possible to collect complete information for each person. Given the extensive resources devoted to the project, every effort should be made to use whatever data were collected.

METHOD

The National Academy of Sciences Committee on the Performance of Military Personnel, an oversight committee for the Joint Service JPM Project, recommended using an imputation procedure that estimates missing data so that complete-case analysis can be conducted [4]. The recommended imputation algorithm is a regression-based procedure that seeks to impute missing values by taking into account the differing levels of task difficulties while maintaining individual differences among examinees [5]. The procedure incorporates a random component equal to the standard error of the estimate to prevent unduly high correlations among variables with imputed values, as compared with variables with nonimputed values. The procedure also sequentially estimates multiple missing values for the same person using a multistage process that relies on previously imputed values for the imputation of successive missing values. Further discussion of the computational procedures for data imputation is presented in appendix B.

Hands-on performance data were collected at the step level; examinees either passed or failed. Data were missing at the step level for a number of reasons:

- Equipment necessary for completion of the step was missing. This happened at one site for a day because equipment was needed for Operation Desert Shield.
- Broken equipment prevented step completion. Items such as oil filters and plugs occasionally broke without an available replacement.
- The test administrator's view was blocked, preventing observation of the step performed.
- The examinee was called away from testing before completing a task.
- The examinee refused to perform a task for personal safety reasons.
- The test site was temporarily inoperable because of weather conditions.

Table 2 details the number of tasks and steps for each performance test. Because of the large number of steps, there were many possibilities for missing data. Rather than exclude a data case with some missing data, step scores were imputed as required to obtain a complete record for that person.

Table 2. Number of steps in performance tests, by MOS

MOS	Hands-on tasks		Manuals task	Forms task
	Tasks	Steps	Steps	Steps
6112	19	332	16	23
6113	16	241	17	26
6114	15	203	16	23
6115	16	240	15	26

GAINS IN COMPLETE CASES

Tables 3 through 6 detail the gains in complete-data cases resulting from imputation of missing data, by task. Overall, complete data were collected for over 95 percent of all tasks administered. Across the four MOSs, with the exception of three tasks in MOS 6112, all tasks had more than 85 percent of the cases with a full complement of steps, without need for any imputation whatsoever.

Difficulties were greater for MOS 6112 because it had the largest number of steps. There were many missing cases at site B, where it was the first MOS to be tested. For MOS 6112, three tasks had less than 85 percent complete cases: task 6A (remove/replace unitary pedal adjustment, 62.1 percent), task 8B (trouble-shoot power plant oil system, 63.8 percent), and task 8C (remove/replace power plant fuel boost pump, 48.3 percent). The missing data for these tasks were mostly the result of difficulties keeping a supply of spare parts.

For task 8B, there were problems maintaining straws for extracting oil; for task 8C, it was difficult keeping an adequate supply of caps. Task 6A had steps missing for a different reason: At site B, it was decided in the middle of testing to perform step 7 before the testing procedure. Therefore, most site B participants had step 7 missing. If these particular steps were ignored, MOS 6112 would have a very high percentage of complete data. There would be 96 percent complete data for task 6A, 97.7 percent complete data for task 8B, and 99.4 percent complete data for task 8C.

Table 3. Complete-step cases by task for MOS 6112 (total n = 174)

Task no.	Task name	Total steps	Number of complete-step cases	Percentage of complete-step cases
1A	Remove/replace power train drive shaft	17	167	96.0
1B	Troubleshoot main gear box drive shaft	8	144	82.8
2A	Troubleshoot power train main gear box oil system	13	173	99.4
2B	Troubleshoot main gear box chip lights	21	171	98.3
2C	Remove/replace power train sump oil filter	21	172	98.9
2D	Ground handling	22	173	99.4
3A	Troubleshoot flight controls	18	170	97.7
3B	Troubleshoot power plant fuel system	32	166	95.4
3C	Turn off/on auxiliary power unit	25	171	98.3
4A	Remove/replace main rotor head	29	170	97.7
4B	Troubleshoot rotor head	12	152	87.4
5A	Remove/replace rotating scissors drive collar	26	152	87.4
5B	Adjust pitch control rod	9	171	98.3
6A	Remove/replace unitary pedal adjustment	17	108	62.1[a]
7A	Remove/replace collective pitch bellcrank	16	169	97.1
7B	Adjust flight control rod	5	169	97.1
8A	Troubleshoot power plant control system	8	172	98.9
8B	Troubleshoot power plant oil system	11	111	63.8[b]
8C	Remove/replace power plant fuel boost pump	22	84	48.3[c]
Form	Forms	23	172	98.9
Manual	Manuals	16	174	100.0

a. For task 6A, one-third of the cases were missing step 7, "remove cotter pin, nut, washer, and bolt attaching clevis to mechanism," because it was decided at Tustin to perform the step in setup, before the actual testing procedure. If that step were ignored, there would be 96 percent complete data for task 6A.

b. For task 8B, Tustin had no straws for extracting oil, so one-third of the total could not perform this task. Ignoring those who were missing the straw, there was 97.7 percent complete data for task 8B.

c. For task 8C, Tustin lacked caps, which resulted in 49.4 percent of the total having two missing steps. Ignoring the steps that required caps, there was 99.4 percent complete data for task 8C.

Appendixes C through F present the degree of imputation for the cases that had any missing steps. For MOS 6112, almost all cases (95.4 percent) had imputation for 4 or fewer tasks out of 19. For those examinees with missing steps, an average of 11 steps out of 332 (3.3 percent) were imputed.

Table 4. Complete-step cases by task for MOS 6113 (total n = 120)

Task no.	Task name	Total steps	Number of complete-step cases	Percentage of complete-step cases
1B	Troubleshoot main gear box drive shaft	8	149	99.2
1B	Troubleshoot flight controls	11	119	99.2
1C	Remove/replace power train sump oil filter	21	120	100.0
1D	Ground handling and taxiing	22	120	100.0
2A	Remove/replace power train drive shaft	11	118	98.3
3A	Adjust pitch control rod	10	119	99.2
3B	Remove/replace rotating scissors and shim	13	119	99.2
4A	Troubleshoot power train main gear box drive shaft	6	120	100.0
4B	Remove/replace mechanical screwjack	12	120	100.0
5A	Remove/replace collective pitch bellcrank	16	120	100.0
5B	Remove/replace power train oil filter	21	118	98.3
6A	Troubleshoot rotor head	18	111	92.5
6B	Remove/replace main rotor head	18	120	100.0
7A	Troubleshoot power train main gear box—chip lights	14	113	94.2
7B	Troubleshoot power plant oil system	21	119	99.2
8A	Remove/replace fuel boost pump	31	120	100.0
Form	Forms	26	116	96.7
Manual	Manuals	17	115	95.8

Table 5. Complete-step cases by task for MOS 6114 (total n = 215)

Task no.	Task name	Total steps	Number of complete-step cases	Percentage of complete-step cases
1A	Remove/replace unitary pedal adjuster	14	210	97.7
1B	Ground handling and taxiing	22	203	94.4
2A	Remove/replace collective pitch bellcrank	13	211	98.1
3A	Troubleshoot main gear box chip detectors	18	212	98.6
3B	Remove/replace power train oil filter	12	212	98.6
4A	Remove/replace power train main gear box drive shaft	17	210	97.7
5A	Troubleshoot rotor head	13	209	97.2
5B	Remove/replace drive link	17	199	92.6
6A	Remove/replace main rotor head	22	212	98.6
6B	Adjust flight control rod	5	213	99.1
7A	Troubleshoot flight controls	14	207	96.3
7B	Troubleshoot power plant oil system	9	207	96.3
8A	Remove/replace power plant fuel boost pump	12	215	100.0
8B	Troubleshoot power plant fuel system	9	183	85.1
8C	Troubleshoot power plant control system	6	198	92.1
Form	Forms	23	212	98.6
Manual	Manuals	16	215	100.0

Table 6. Complete-step cases by task for MOS 6115 (total n = 149)

Task no.	Task name	Total steps	Number of complete-step cases	Percentage of complete-step cases
1A	Turn off/on auxiliary power unit	26	149	100.0
1B	Troubleshoot flight controls	15	148	99.3
1C	Adjust flight control rod	5	149	100.0
1D	Ground handling and taxiing	22	149	100.0
2A	Remove/replace accessory gear box drive shaft	10	146	98.0
3A	Adjust pitch control rod	9	147	98.7
3B	Remove/replace rotating scissors and shims	13	147	98.7
4A	Remove/replace mechanical screwjack	15	146	98.0
4B	Troubleshoot power plant fuel system	12	149	100.0
5A	Remove/replace collective pitch bellcrank	16	145	97.3
6A	Troubleshoot main rotor head	22	148	99.3
6B	Remove/replace main rotor	16	135	90.6
7A	Troubleshoot main gear box—chip detectors	15	148	99.3
7B	Remove/replace power train accessory gear box oil filter	19	149	100.0
7C	Troubleshoot power plant oil system	7	147	98.7
8A	Remove/replace power plant fuel boost pump	18	145	97.3
Form	Forms	26	145	97.3
Manual	Manuals	15	148	99.3

Less imputation was necessary for MOS 6113: 82 percent of cases needed no imputation whatsoever, and 91.7 percent of cases would be complete with imputation of one or no steps. All examinees had three or fewer tasks imputed. For the 22 cases with any imputation, an average of 3 steps out of 241 (1.2 percent) were imputed to achieve complete task-level information. These statistics show a very small amount of missing data for MOS 6113.

Little imputation was necessary for MOS 6114: 67 percent of cases needed no imputation whatsoever, and almost all cases (99 percent) required imputation of steps in 4 or fewer tasks out of 15. The average number of imputed steps for those who had imputed steps was 17 out of a total of 203 (8.4 percent).

MOS 6115 had a high percentage of complete data: 155 examinees (77.2 percent) needed no imputation whatsoever, and all examinees required imputation of two or fewer tasks. For those who required any imputation, an average of 3 steps out of 240 (1.3 percent) were imputed.

CHANGES IN MEANS AND STANDARD DEVIATIONS

Given this degree of imputation at the step level, what was the impact on the sample statistics of the respective hands-on scores? Tables 7 through 10 show the changes in mean and standard deviation caused by imputation. With the exception of MOS 6112, the shifts in mean performance are relatively small compared to the standard deviation of the performance scores. The shift in mean performance for MOS 6112 was attributable to the fact that all complete cases were at site A, which generally had higher scores than at site B for all MOSs. The difference in scores by site was the result of artifactual scoring differences between the two sites [6]. The site was included in the imputation procedure as a predictor variable so that across-site scoring differences were maintained. Other analyses examined the cause of these scoring differences and made appropriate adjustments [6]. Table 11 shows the differences in total score by site for complete cases. Because site was used as a predictor for imputation, imputation tended to bring scores lower. This was especially true for MOS 6112, where the imputed cases were much more likely to be from site B than from site A.

Table 7. Comparison of task statistics using original and imputed data (MOS 6112)

	Original data					Imputed data				
Task	n	Mean	Std. dev.	Min.	Max.	n	Mean	Std. dev.	Min.	Max.
1A	167	85.8	13.1	44	100	174	85.8	13.0	44	100
1B	144	75.1	22.3	13	100	174	74.0	22.6	13	100
2A	173	87.8	19.4	19	100	174	87.9	19.3	19	100
2B	171	76.7	23.2	0	100	174	77.1	23.1	0	100
2C	172	86.9	14.1	48	100	174	86.9	14.1	48	100
2D	173	78.3	16.7	0	100	174	78.3	16.6	0	100
3A	170	81.4	19.3	24	100	174	81.4	19.2	24	100
3B	166	93.0	12.5	33	100	174	92.9	12.5	33	100
3C	171	91.5	12.4	40	100	174	91.4	12.4	40	100
4A	170	90.6	10.4	54	100	174	90.5	10.5	54	100
4B	152	67.9	25.9	6	100	174	66.7	25.5	6	100
5A	152	92.2	8.1	54	100	174	91.9	8.2	54	100
5B	171	82.0	14.3	33	100	174	82.0	14.2	33	100
6A	108	88.5	16.7	19	100	174	87.0	19.1	13	100
7A	169	85.1	16.7	7	100	174	85.2	16.5	7	100
7B	169	86.2	18.1	20	100	174	84.6	20.9	0	100
8A	172	89.8	12.2	50	100	174	89.4	12.6	50	100
8B	111	89.3	12.1	40	100	174	91.8	12.4	32	100
8C	84	85.4	10.5	56	100	174	85.4	11.7	50	100
Forms	172	65.7	23.7	0	100	174	65.5	23.6	0	100
Manuals	174	83.6	14.5	31	100	174	83.6	14.5	31	100
PCTOT[a]	48	88.6	5.4	73	97	174	83.7	7.9	63	97

NOTE: Imputed data include both original complete cases and cases that required some imputation.

a. PCTOT is total score on all tasks.

Table 8. Comparison of task statistics using original and imputed data (MOS 6113)

Task	Original data					Imputed data				
	n	Mean	Std. dev.	Min.	Max.	n	Mean	Std. dev.	Min.	Max.
1A	116	80.6	15.9	22	100	117	80.6	15.9	22	100
1B	116	85.3	19.0	0	100	117	85.5	19.0	0	100
1C	117	86.7	20.3	20	100	117	86.7	20.3	20	100
1D	117	70.9	18.0	16	100	117	70.9	18.0	16	100
2A	115	82.1	16.4	27	100	117	81.9	16.3	27	100
3A	116	79.1	16.3	20	100	117	78.8	16.5	20	100
3B	116	72.6	20.0	0	100	117	72.5	20.0	0	100
4A	117	93.4	15.8	17	100	117	93.4	15.8	17	100
4B	117	88.1	10.7	58	100	117	88.1	10.7	58	100
5A	117	95.8	8.8	50	100	117	95.8	8.8	50	100
5B	115	90.4	12.9	0	100	117	89.7	15.3	0	100
6A	109	69.4	27.8	0	100	117	67.8	28.5	0	100
6B	117	74.6	21.4	0	100	117	74.6	21.4	0	100
7A	110	78.3	20.0	0	100	117	77.9	21.2	0	100
7B	116	78.0	15.4	25	100	117	78.0	15.4	25	100
8A	117	89.3	12.4	38	100	117	89.3	12.4	38	100
Forms	116	53.6	20.5	0	100	117	53.6	20.5	0	100
Manuals	115	81.8	19.1	13	100	117	81.6	19.1	13	100
PCTOT[a]	96	80.6	8.5	52	93	117	80.4	8.7	52	94

NOTE: Where n = 117, three cases were deleted because of aberrant scores.

a. PCTOT is total score on all tasks.

Table 9. Comparison of task statistics using original and imputed data (MOS 6114)

Task	Original data					Imputed data				
	n	Mean	Std. dev.	Min.	Max.	n	Mean	Std. dev.	Min.	Max.
1A	210	89.1	12.5	33	100	215	89.1	12.4	33	100
1B	203	66.5	13.9	35	100	215	65.9	13.9	35	95
2A	211	94.1	13.5	34	100	215	94.1	13.4	34	100
3A	212	64.9	32.4	9	100	215	64.6	32.2	9	100
3B	212	84.2	19.7	0	100	215	84.3	19.7	0	100
4A	210	76.2	20.9	0	100	215	76.2	21.0	0	100
5A	209	78.8	19.6	15	100	215	78.8	19.5	15	100
5B	199	49.2	29.9	0	100	215	49.0	29.7	0	100
6A	212	87.1	9.4	59	100	215	86.8	9.7	56	100
6B	213	80.9	20.9	25	100	215	80.7	20.9	25	100
7A	207	65.5	20.2	15	100	215	66.0	20.1	15	100
7B	207	89.3	12.1	33	100	215	89.2	12.1	33	100
8A	215	76.7	21.2	20	100	215	76.7	21.2	20	100
8B	183	76.4	21.7	22	100	215	75.1	21.1	22	100
8C	198	70.5	26.6	17	100	215	69.9	26.4	17	100
Forms	212	62.4	20.6	9	100	215	62.3	20.5	9	100
Manuals	215	85.1	16.4	15	100	215	85.1	16.4	15	100
PCTOT[a]	143	77.8	8.5	55	96	215	76.1	9.5	50	96

a. PCTOT is total score on all tasks.

Table 10. Comparison of task statistics using original and imputed data (MOS 6115)

Task	Original data					Imputed data				
	n	Mean	Std. dev.	Min.	Max.	n	Mean	Std. dev.	Min.	Max.
1A	149	81.2	17.3	18	100	149	81.2	17.3	18	100
1B	148	68.0	27.1	0	100	149	67.6	27.4	0	100
1C	149	91.5	15.1	40	100	149	91.5	15.1	40	100
1D	149	77.0	15.5	25	100	149	77.0	15.5	25	100
2A	146	92.9	10.9	40	100	149	92.8	10.8	40	100
3A	147	91.8	14.6	0	100	149	91.4	15.0	0	100
3B	147	87.2	14.4	0	100	149	87.1	14.4	0	100
4A	146	85.6	15.1	43	100	149	85.8	15.0	43	100
4B	149	83.9	21.7	9	100	149	83.9	21.7	9	100
5A	145	92.5	11.5	25	100	149	92.5	11.3	25	100
6A	148	70.8	28.0	5	100	149	70.9	27.9	5	100
6B	135	83.9	18.8	0	100	149	83.0	18.4	0	100
7A	148	93.3	9.2	51	100	149	93.3	9.2	51	100
7B	149	81.4	17.7	42	100	149	81.4	17.7	42	100
7C	147	70.3	23.1	14	100	149	70.5	23.0	14	100
8A	145	94.2	8.8	43	100	149	94.2	8.7	43	100
Forms	145	61.6	18.7	0	100	149	61.3	18.8	0	100
Manuals	148	86.9	16.1	29	100	149	87.0	16.1	29	100
PCTOT[a]	115	83.9	9.9	45	98	149	82.9	9.7	45	98

a. PCTOT is total score on all tasks.

Table 11. Comparison of total scores for complete cases, by site

MOS	Site A			Site B		
	n	Mean	Std. dev.	n	Mean	Std. dev.
6112	42	87.9	5.3	–	–	–
6113	–	–	–	69	79.9	8.8
6114	35	84.4	6.5	83	74.1	6.9
6115	87	90.1	5.1	46	77.4	8.9

NOTE: These n's do not match tables 7 through 10 because listwise deletion was used. There were no complete cases for site B, MOS 6112. All testing for MOS 6113 was conducted at site B.

COMPARISON OF CORRELATIONS BY TASK

In almost all cases, correlations changed very little by task as a result of imputation, as shown in appendixes G through J. Because imputation was done by task, these tables show that imputation did not result in particular tasks becoming more (or less) correlated with aptitude. The one exception to the rule was for MOS 6112 tasks 6A, 8B, and 8C. In these cases, when listwise deletion[1] was used, there were extremely few cases at site B for the particular task (15 complete cases for 6A, 16 complete cases for 8B, and 2 complete cases for 8C). These anomalies occurred because there were several steps for which nearly all people had missing data at site B. In cases involving very few complete data, imputation could have a large influence on correlations because the original correlations were based on too small a sample. The final correlations for these tasks are reasonable, however, given that the final correlations lie within the range of values for the other tasks, many of which had mostly complete cases.

COMPARISON OF TOTAL SCORE CORRELATIONS OVERALL

Table 12 compares correlations before and after imputation by site and overall, within MOS. For MOS 6113, the only MOS for which testing occurred at only one site, the correlations of aptitude and total hands-on score decreased very slightly. MOS 6113 shows that the imputation program functioned as it should, preserving the general strength of correlations among variables of interest. The very slight decrease in correlation with aptitude occurs because the imputation procedure incorporates a random variable component to reflect the uncertainty of task- and step-level predictions. There was a slight increase in correlation with time in service (TIS) as a result of imputation because both TIS and TIS^2 were used to impute scores.

For the three MOSs (6112, 6114, and 6115) that had separate testing sites, the data that included the imputed hands-on scores had a slightly increased correlation with enlistment MM across sites. However, the overall correlation with enlistment aptitude was within or lower than the correlations for the two sites separately. For MOS 6112, the separate site correlations before imputation were .35 and .43, but the overall imputed correlation was .28. Similarly, the separate site correlations for MOS 6114 were .31 and .33, and the overall imputed correlation was .33. For MOS 6115, the separate site correlations before imputation were .26 and .10, but the overall imputed correlation was .19. These findings indicate that the increases in correlation across sites were reasonable because the correlations remained within the range of those observed separately at the two sites.

1. Listwise deletion means that only cases with complete data on all variables were used.

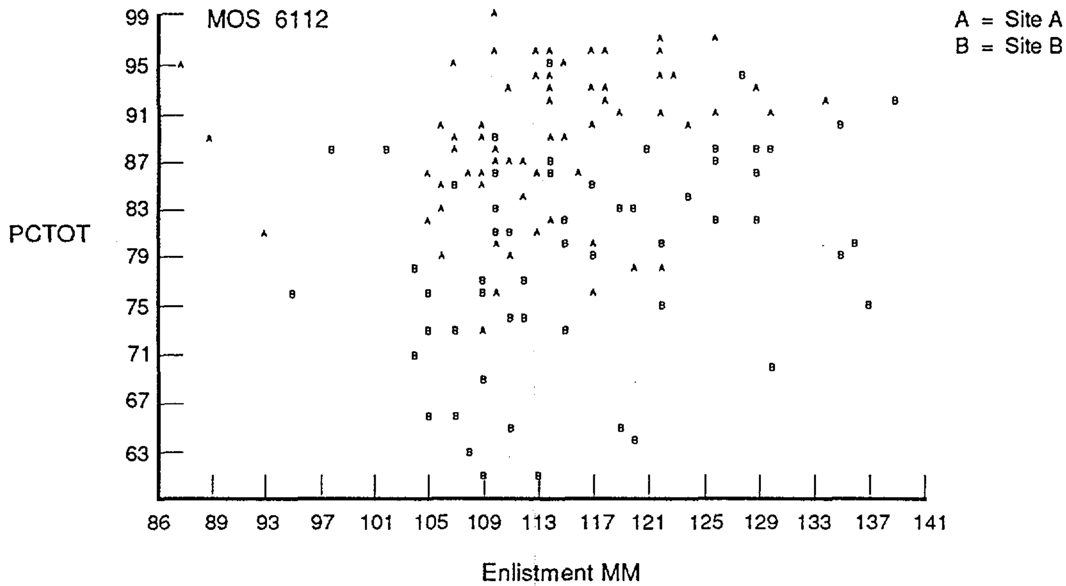
Table 12. Correlation of enlistment MM, CAT-MM, and TIS with PCTOT by site and MOS

	Enlistment MM				CAT-MM				TIS			
	Before imputes		After imputes		Before imputes		After imputes		Before imputes		After imputes	
MOS	n	r	n	r	n	r	n	r	n	r	n	r
6112[a]												
Site A	65	.35	85	.38	65	.52	85	.49	65	.63	85	.70
Site B	54	.43	66	.48	54	.47	66	.48	54	.48	66	.41
All	119	.22	151	.28	119	.31	151	.36	119	.45	151	.54
6113												
Site B	71	.32	88	.30	71	.55	88	.53	71	.56	88	.61
6114												
Site A	35	.31	49	.21	35	.41	49	.31	35	.00	49	.13
Site B	83	.33	130	.35	83	.38	130	.48	83	.41	130	.42
All	118	.28	179	.33	118	.35	179	.47	118	.33	179	.34
6115												
Site A	41	.26	48	.31	41	.38	48	.41	41	.63	48	.62
Site B	46	.10	66	.22	46	.17	66	.22	46	.39	66	.24
All	87	.12	114	.19	87	.17	114	.18	87	.40	114	.32

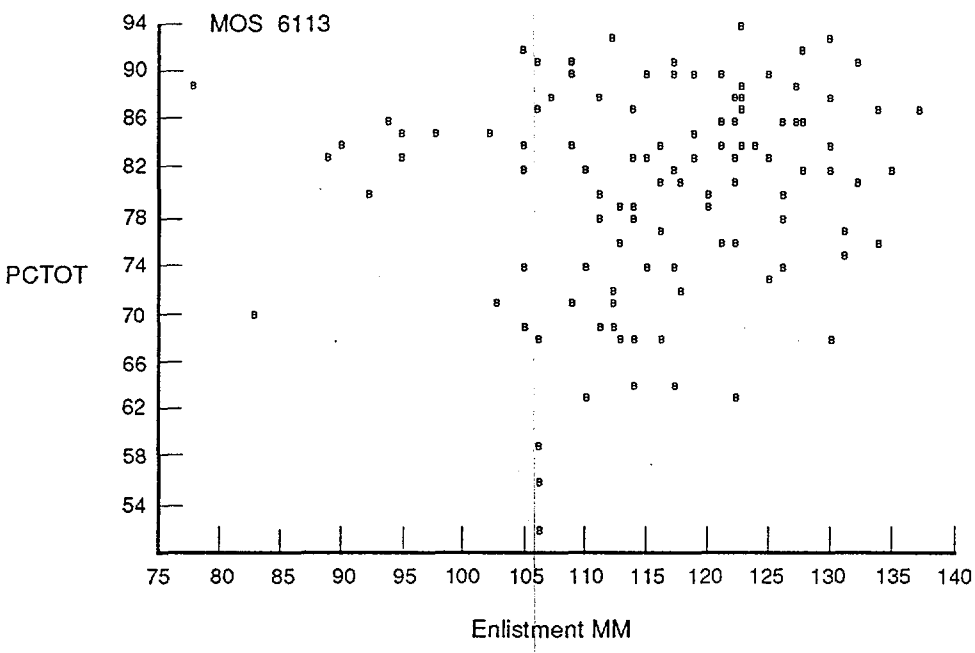
NOTE: r is the correlation coefficient.

a. Numbers for MOS 6112 are based on 13 of the 19 tasks—tasks 1B, 4B, 5A, 6A, 8B, and 8C were left out because of the high percentage of missing data at site B for those tasks. When these six tasks were included, nearly all Marines from site B would have at least some imputed data, and pre-imputation correlations for site B would not be meaningful. Listwise deletion was used.

Table 12 shows that the overall correlations between aptitude and total hands-on score were lowered by the fact that site A's scores were higher than site B's. Putting data from the two sites together usually reduced the overall correlations to below the average of the correlations for either site separately. For example, whereas the separate site correlations before imputation for MOS 6112 with enlistment MM were .35 for site A and .43 for site B, the overall correlation was .22. In similar fashion, the separate site correlations for MOS 6112 CAT-MM were .52 and .47, but the overall correlation was just .31. As can be seen from table 12, this pattern remained even after imputation. Figure 3 shows how combining two sets of scores, each with positive correlations, reduced overall correlations between aptitude and hands-on total scores. For MOS 6113 (panel B), testing occurred only at site B.

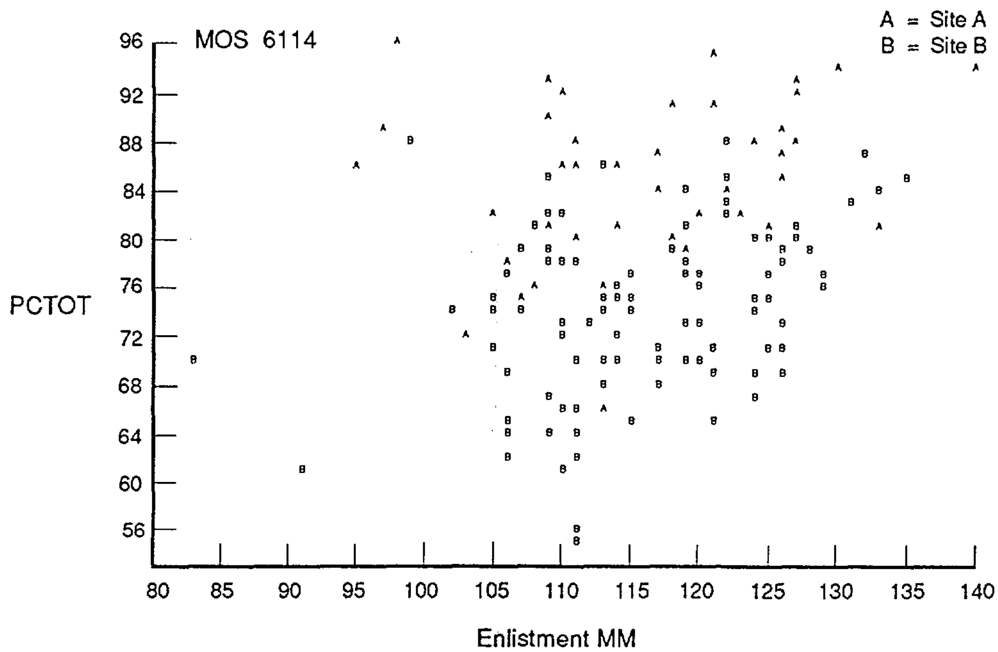


NOTE: 3 observations had missing values; 11 were hidden.

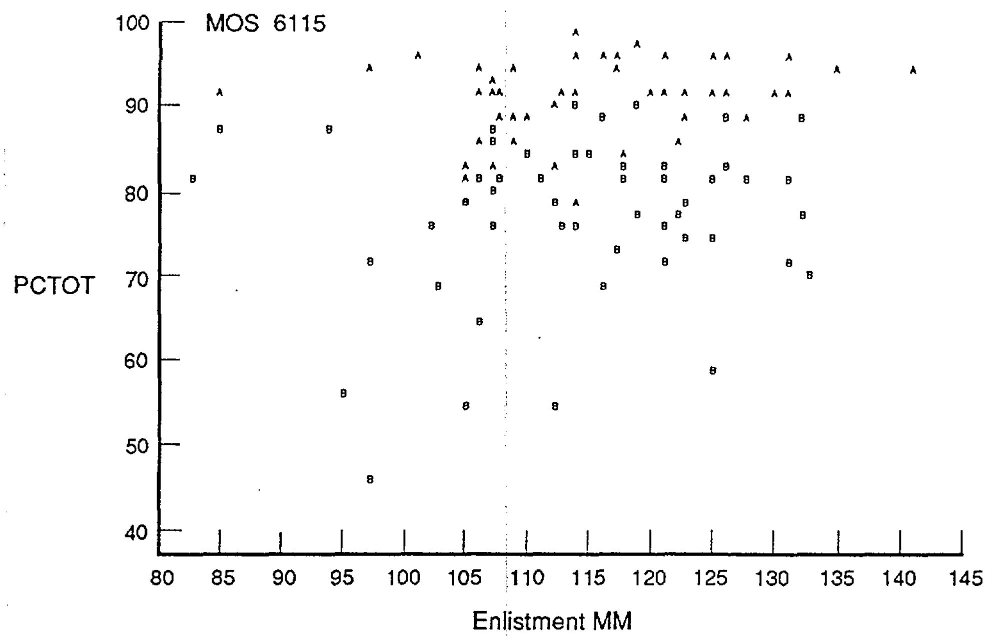


NOTE: 8 observations had missing values; 5 were hidden.

Figure 3. Effects of combining data from two sites on hands-on performance/aptitude relationships



NOTE: 3 observations had missing values; 7 were hidden.



NOTE: 5 observations had missing values; 14 were hidden.

Figure 3. (Continued)

COMPARISON OF CORRELATIONS WITHIN SITE

For the MOSs that had two testing sites (MOS 6112, 6114, 6115), correlations with aptitude also tended to increase somewhat within site (table 12). The four plausible causes of the increase within site follow:

1. It was an artifact of imputation adding cases that were already more extreme.

2. It was an artifact of different correlation patterns between sites and between complete and imputed-only cases.

3. It was an artifact of increased reliability as a result of imputation.

4. The imputation program was functioning improperly.

The first three causes are partially responsible for the increase in correlation—the fourth explanation can be shown to be false.

Table 13 demonstrates that the imputation process was not the reason for correlational increase. It shows the correlations that would be computed if, rather than imputing missing data, each person's average score across all tasks were used to replace scores for tasks that had any imputed steps. As can be seen, substitution of average scores resulted in nearly the same correlations as did the imputation procedure. This finding demonstrates that the imputation process did not increase correlation per se and that increases in correlation with imputation were substantially a result of adding cases that were already more extreme. Cause 1, therefore, was substantially responsible for the correlation increase.

Cause 2—differences in correlation patterns—also contributed to the increase in correlations. Tables 14 and 15 show that the two sites differed in their correlation patterns and time in service. Complete cases for site A generally had more TIS than did complete cases for site B (table 14). For MOS 6114, complete cases for site A had slightly higher aptitude than did site B. Because TIS was correlated with success on hands-on tasks and there were more low-TIS cases added as a result of imputation, imputation brought scores lower.

Table 15 indicates that differences in correlation patterns between sites contributed to the correlational increase. In two cases, increases in correlation were artifacts of the correlation between TIS and aptitude. This occurred for site A in

MOS 6112 and MOS 6115 imputed-only data, where TIS correlated with enlistment MM (.33 and .46). An artifact also contributed to changes in correlation for MOS 6114. For MOS 6114, the correlation between TIS and total score for complete cases was .41 for site B but .00 at site A.

Table 13. Comparison of imputation with insertion of average value (correlation with total hands-on score)

	Enlistment MM			CAT-MM			TIS		
	n	r		n	r		n	r	
MOS		Imp.	Average		Imp.	Average		Imp.	Average
6112[a]									
Site A	85	.38	.37	85	.49	.48	85	.70	.70
Site B	66	.48	.49	66	.48	.50	66	.41	.43
Overall	151	.28	.28	151	.36	.33	151	.54	.48
6113									
Site B	88	.30	.30	88	.53	.53	88	.61	.60
6114									
Site A	49	.21	.21	49	.31	.31	49	.13	.13
Site B	130	.35	.37	130	.48	.49	130	.42	.40
Overall	179	.33	.34	179	.47	.47	179	.34	.33
6115									
Site A	48	.31	.30	48	.41	.40	48	.62	.62
Site B	66	.22	.22	66	.22	.23	66	.24	.22
Overall	114	.19	.19	114	.18	.18	114	.32	.31

NOTE: r is the correlation coefficient. Average values were the average scores for that person across all tasks—the average column shows correlations obtained when average values were used for tasks that would otherwise have been imputed. Listwise deletion was used.

a. For this table, the same 13 MOS 6112 tasks were used as were used in table 12. If all 19 tasks were used, within-site correlations for site B would be based on too small a sample.

Cause 3 was that increased reliability enhanced correlations. This was a minor factor in the increased correlations. As a result of imputation, Cronbach alpha increased from .78 to .82 for MOS 6112, remained at .83 for MOS 6113, increased from .74 to .78 for MOS 6114, and decreased from .87 to .86 for MOS 6115. Because the correction for attenuation due to unreliability is proportional to the square of the reliability, an increase in reliability of .04 will increase the correlation from .01 to .02 if the actual correlation between variables were in the range of .30 to .50.

Table 14. Descriptive statistics for time in service of original and imputed-only data, by site

	Original data			Imputed-only data		
MOS	Mean	Std. dev.	n	Mean	Std. dev.	n
6112						
Site A	47.1	23.0	42	36.1	24.5	43
Site B	–	–	–	40.8	25.3	66
6113						
Site A	–	–	–	–	–	–
Site B	34.7	19.0	69	35.6	26.4	16
6114						
Site A	51.9	31.5	35	26.9	16.4	14
Site B	39.1	20.2	83	32.8	19.0	47
6115						
Site A	52.8	27.3	41	42.6	11.6	7
Site B	48.4	23.9	46	44.3	21.1	20

NOTE: Imputed-only data do not include cases that were initially complete. Original data include only complete cases. Listwise deletion was used.

As a final check on the imputation procedure, plots were made of hands-on total score and enlistment aptitude. Figure 4 shows that imputed cases fell in the full range of both aptitude and performance.

Table 15. Correlation patterns of original and imputed-only data, by site

MOS	Original data				Imputed-only data			
	TIS[a]	ENL-MM[b]	CAT-MM[c]	PCTOT[d]	TIS	ENL-MM	CAT-MM	PCTOT
6112								
Site A								
TIS	1.00	.14	.39	.52	1.00	.33	.47	.84
ENL-MM	.14	1.00	.68	.15	.33	1.00	.66	.36
CAT-MM	.39	.68	1.00	.40	.47	.66	1.00	.50
PCTOT	.52	.15	.40	1.00	.84	.36	.50	1.00
Site B								
TIS					1.00	.04	.17	.45
ENL-MM					.04	1.00	.80	.45
CAT-MM					.17	.80	1.00	.47
PCTOT					.45	.45	.47	1.00
6113								
Site B								
TIS	1.00	.11	.41	.59	1.00	.08	.18	.76
ENL-MM	.11	1.00	.65	.33	.08	1.00	.60	.41
CAT-MM	.41	.65	1.00	.58	.18	.60	1.00	.41
PCTOT	.59	.33	.58	1.00	.76	.41	.41	1.00
6114								
Site A								
TIS	1.00	.04	.27	.00	1.00	-.06	.15	.57
ENL-MM	.04	1.00	.67	.31	-.06	1.00	.70	.08
CAT-MM	.27	.67	1.00	.41	.15	.70	1.00	.22
PCTOT	.00	.31	.41	1.00	.57	.08	.22	1.00
Site B								
TIS	1.00	.11	.23	.41	1.00	-.18	.21	.37
ENL-MM	.11	1.00	.65	.33	-.18	1.00	.64	.40
CAT-MM	.23	.65	1.00	.38	.21	.64	1.00	.54
PCTOT	.41	.33	.38	1.00	.37	.40	.54	1.00
6115								
Site A								
TIS	1.00	.07	-.09	.63	1.00	.46	.68	.71
ENL-MM	.07	1.00	.67	.26	.46	1.00	.65	.77
CAT-MM	-.09	.67	1.00	.38	.68	.65	1.00	.77
PCTOT	.63	.26	.38	1.00	.71	.77	.77	1.00
Site B								
TIS	1.00	.02	.18	.39	1.00	-.03	.08	-.42
ENL-MM	.02	1.00	.75	.10	-.03	1.00	.71	.64
CAT-MM	.18	.75	1.00	.17	.08	.71	1.00	.47
PCTOT	.39	.10	.17	1.00	-.42	.64	.47	1.00

NOTE: There were no complete cases at site B for MOS 6112. n's for MOS 6112 were 42 and 43 for site A, 0 and 66 for site B for original and imputed-only data, respectively. Corresponding n's for MOS 6113 were 69 and 16. For MOS 6114, n's were 35 and 14 for site A, 83 and 47 for site B. For MOS 6115, n's were 41 and 7 for site A; 46 and 20 for site B, respectively. n's are for listwise deletion.

a. TIS = time in service.
b. ENL-MM = enlistment ASVAB mechanical maintenance aptitude.
c. CAT-MM = computer adaptive testing ASVAB mechanical maintenance aptitude.
d. PCTOT = total score on all tasks.

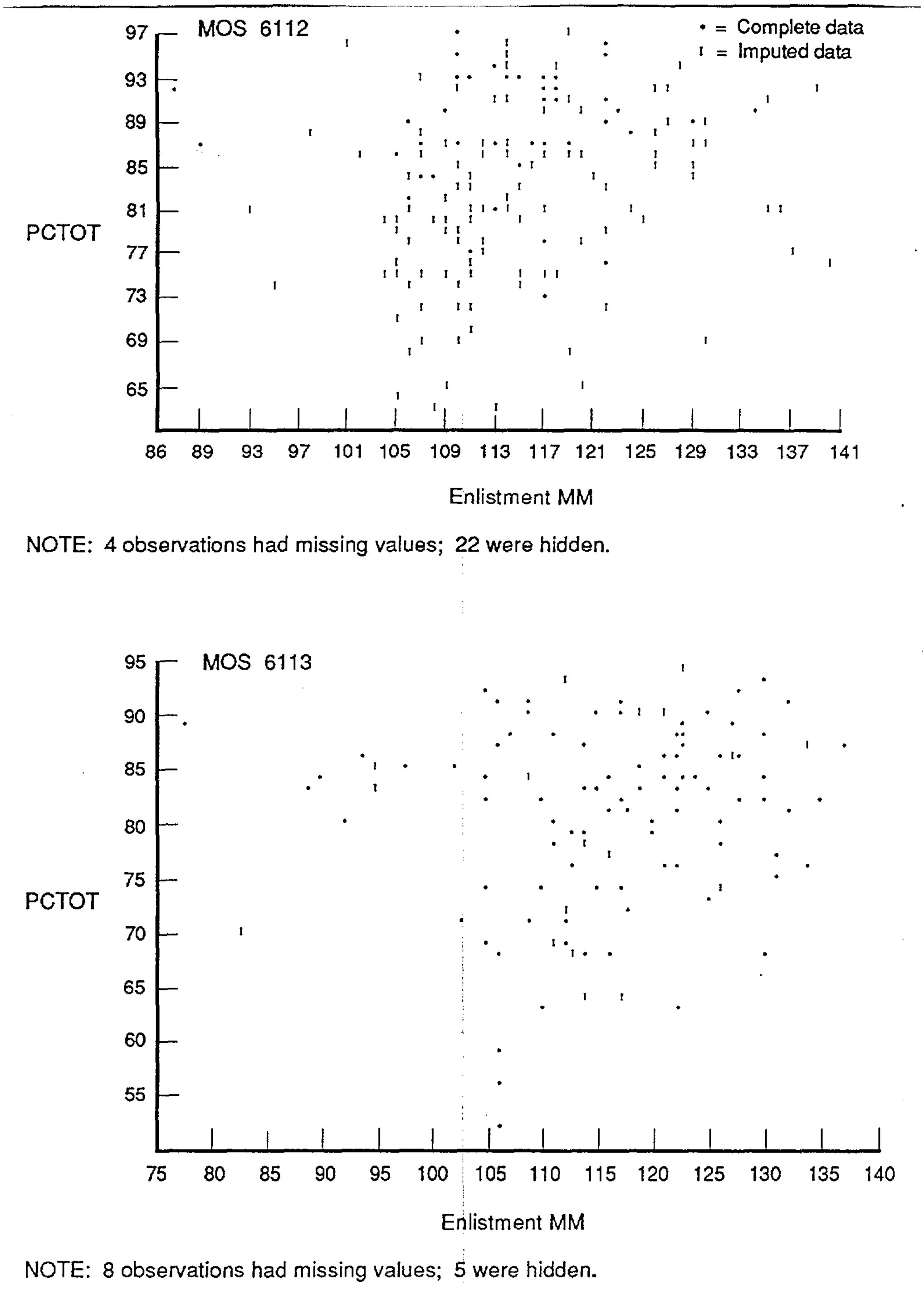
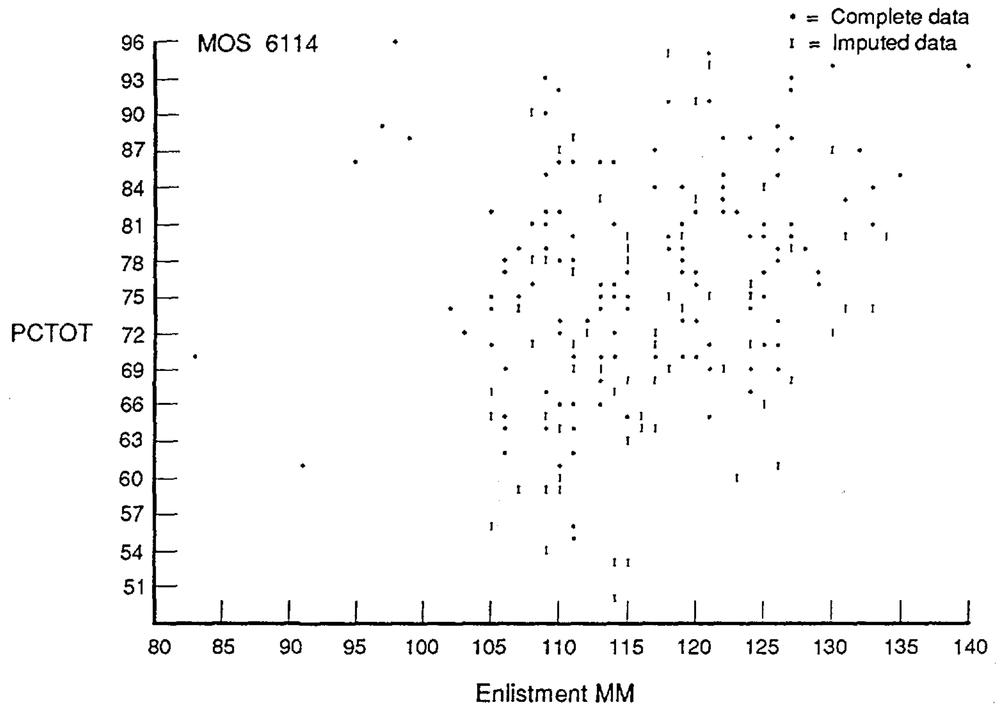
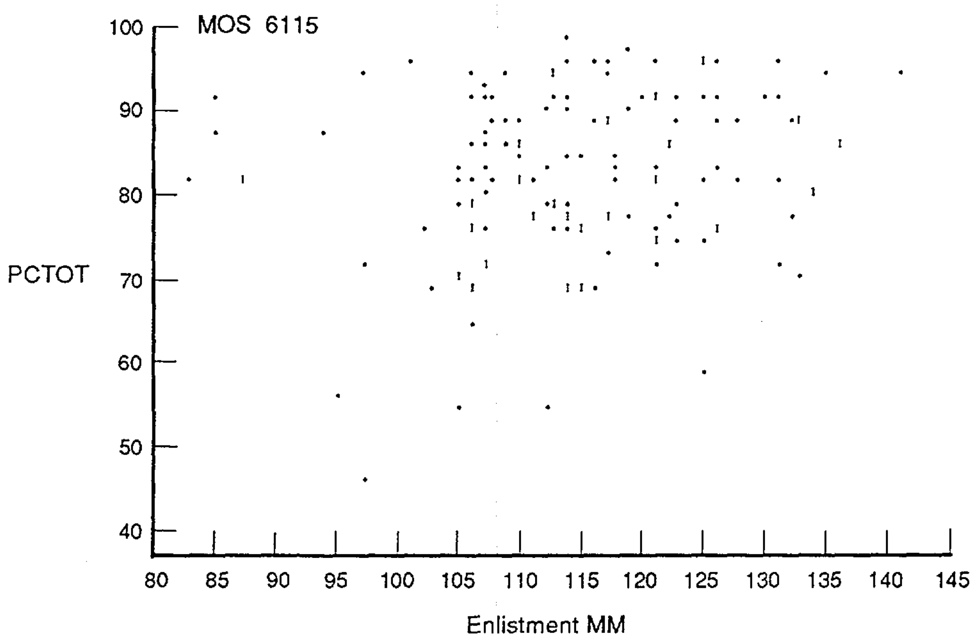


Figure 4. Scatterplots of hands-on total score versus mechanical maintenance composite score for both imputed and complete data



NOTE: 3 observations had missing values; 22 were hidden.



NOTE: 8 observations had missing values; 23 were hidden.

Figure 4. (Continued)

REPORTED DATA COLLECTION PROBLEMS

As described earlier, there were several data collections in addition to hands-on mechanical testing. The ECAT was a computerized test battery of new predictors of job performance, such as one-hand tracking, short-term memory, and reaction time. The CAT-ASVAB is a computerized adaptive version of the regular ASVAB. The "administrative duties" test was a hands-on test of the Marine's proficiency in two tasks—looking up information in manuals and filling out forms. Three aptitudes were tested by the General Aptitude Test Battery (GATB) portions taken: motor coordination, manual dexterity, and finger coordination. The tasks involved writing as many symbols as possible, placing rivets, or assembling rivets in a constrained time period. The job knowledge test (JKT) consisted of multiple-choice paper-and-pencil items designed to parallel the hands-on tasks as closely as possible.

Relatively few missing data were reported for the CAT-ASVAB, administrative duties, GATB, and JKT data collection, as shown in table 16. Nevertheless, it is useful to pinpoint the reasons for missing data to direct the design of future data collection efforts.

Table 16. Amount of missing data for CAT-ASVAB, administrative duties test, GATB, and JKT

	MOS and site									
	6112		6113	6114		6115		Total		
Test	A	B	B	A	B	A	B	A	B	Overall
CAT-ASVAB	0	1	10	0	12	0	1	0	24	24 (3.6%)
ECAT	0	3	5	1	1	1	10	2	19	21 (3.2%)
GATB	0	1	0	0	1	0	0	0	2	2 (0.3%)
JKT	0	0	0	0	0	0	0	0	0	0 (0.0%)

NOTE: Percentages are based on the entire sample of 658 hands-on cases.

This section analyzes the problems that were reported concerning data collection. This information is important because it describes the difficulties encountered in collecting performance data from multiple sources.

At both sites, test administrators filled out daily logs that described any abnormalities or deviations from expected test procedures for the ECAT, CAT, administrative duties test, GATB, and JKT. The frequency of particular categories of reported problems from these logs forms the basis of the following analyses. Analyses of problem logs/failure reports could be useful for (a) pointing out relative magnitudes of problems within sites, (b) analyzing the degree to which problems are recognized by test administrators, and (c) suggesting the magnitude of problems that might occur if these tests were to be given operationally. For example, depending on policy decisions, some form of the ECAT and CAT-ASVAB could be used operationally as replacements for the currently used paper-and-pencil ASVAB.

RESULTS

Documented testing problems were categorized based on the general duties of the test administrator required for each testing session:

- Site setup
  - Determine the completeness and soundness of testing equipment (e.g., computer).
  - Prepare testing equipment for administration.
- Test administration—Deal with problems during the test.
- Data collection and verification
  - Collect test data (e.g., disk information, answer sheets).
  - Confirm that test data have been collected properly, and, if necessary, collect data again.
  - Send data to a central site, with proper identification attached to test responses.

These test administration tasks were the focal point of the training sessions for test administrators in an attempt to prevent missing data and other test administration problems. These are the points at which it is important for the test administrators to confirm that data have been collected.

Because almost all problems with data collection reported in table 16 concerned computerized test administration (CAT-ASVAB or ECAT), the problem log analysis centered on computerized test administration. Table 17 shows that the largest number of computerized test problems occurred at site B and that the largest numbers of problems were because of the response pedestal or a computer error during testing (e.g., the screen going blank, the computer skipping a question for some unknown reason). Examinee motivation was sometimes a problem with both CAT-ASVAB and ECAT.

Table 17. Reported problems of computerized data collection, by site

	ECAT		CAT		
Problem	Site A	Site B	Site A	Site B	Total
Response pedestal	5	7			12
Skipped test portion	1	0	0	0	1
Power cord	0	5	0	4	9
Boot failure	2	0	0	0	2
Collection failure	0	4	0	2	6
Help key	1	1	0	0	2
Identification problem	0	0	0	1	1
Video/computer error	1	12	3	1	17
Examinee didn't try	3	1	7	1	12
Examinee unable (e.g., tired, sick)	1	1	0	0	2
Scheduling	0	0	1	0	1
Outside disruption	0	1	0	1	2
Total	14	32	11	10	67

These results are remarkable for what was not reported in the problem logs. Although site B lost 24 examinees' CAT data, the problem logs did not report the loss because test administrators at site B did not verify data before sending them. ECAT data also presented a problem. At site B, there was confusion about how often to back up the disks. Although site B backed up weekly, they tried to put up to 18 people's data on the same disk, and the disk's capacity was no more than 10. Another problem at site B was that disks were sent with empty files; these files should have been checked for contents before they were sent.

In summary, the problem logs indicated relatively few difficulties collecting data and maintaining data quality. This result confirms the findings from tabulations of missing data. Nevertheless, the problem logs showed that each data source presented characteristic challenges to field data collection: response pedestal problems periodically hampered the ECAT, and data transfer failures sometimes obstructed both ECAT and CAT-ASVAB data collection.

RECOMMENDATIONS

For future data collection efforts, specific procedures should be developed to alleviate some of the most common difficulties. For CAT-ASVAB and ECAT, test administrators should always verify that data have been transferred to backup disks immediately. Test administrators should be trained to check data disks for capacity limits to avoid data overflow problems. Finally, clean, air-conditioned spaces are recommended for administration of computerized tests and might prevent problems with examinee motivation.

CONCLUSIONS

HOPT

Relatively few data quality problems were found for the HOPT. Across the four MOSs tested, only three examinees, all from MOS 6113, had such unexpectedly low scores that it was necessary to delete their HOPT scores. This amount was less than 0.5 percent of the total hands-on data collected and 2.5 percent of the MOS 6113 data. Few data were missing for the HOPT, with overall 95 percent complete data for tasks. Only MOS 6112 had less than 85 percent complete data for any task. With few exceptions, the effect of imputing the remaining points was minimal, in terms of mean, standard deviation, and correlation with aptitude composites. Imputation and insertion of average score resulted in nearly identical correlations, which indicates that imputation by itself had little effect on data characteristics.

JKT

The number of unusual response patterns for the JKT was relatively low. Across the four MOSs, it appears that 14 of 658 (2.1 percent) examinees recorded the wrong test form on their answer sheets, and these aberrations were detected by scoring the tests with both answer keys. After correcting items for miskeying, 15 of 673 items (2.2 percent) had sufficiently poor properties to be deleted from the four JKTs. Last, only six examinees' JKT scores were low enough compared with expected values to be deleted from further consideration. The effect of deleting these few examinees' scores was to increase the average score slightly, to decrease the standard deviation, and to increase the correlation of JKT with aptitude. The changes in all cases were extremely small.

CAT-ASVAB

Residual analyses and problem logs pinpointed five persons who apparently did not make a full effort on the CAT-ASVAB. These five scored far below what would be expected on the basis of their enlistment MM scores and time in service. The small percentage of unusual CAT-ASVAB scores suggests that the inducement of offering to change Marines' scores of record if they improved their enlistment score was generally successful.

REFERENCES

[1] American Institutes for Research Report AIR-70900-FR 02/91, *Develop and Administer Measures for Mechanical Maintenance Occupational Area, Volume II: Test Administration I,* by Jennifer L. Crafts et al., 15 Jun 1991

[2] CNA Research Memorandum 88-259, *Analysis of Data Quality for the Infantry Phase of the Marine Corps Job Performance Measurement Project*, by Paul W. Mayberry, Mar 1989

[3] M. H. Maier. "On the Need for Quality Control in Validation Research." *Personnel Psychology* 41 (1988): 497-502

[4] B. F. Green and H. Wing, eds. *Analysis of Job Performance Measurement Data: Report of a Workshop*. Washington, DC: National Academy Press, 1988

[5] L. L. Wise and D. McLaughlin. *Guidebook for the Imputation of Missing Data*. Palo Alto, CA: American Institutes for Research, 1980

[6] CNA Research Memorandum 91-242, *Development and Scoring of Hands-On Performance Tests for Mechanical Maintenance Specialties*, by Neil B. Carey and Paul W. Mayberry, forthcoming

APPENDIX A

DETAILS OF RESIDUAL ANALYSES

APPENDIX A

DETAILS OF RESIDUAL ANALYSES[1]

Residual analyses are conducted to determine whether the assumptions underlying linear regression are correct and to check for outlier scores that might not reflect the examinee's true ability.

Figure A-1 shows a relationship between criterion and predictor. Note that the criterion is represented along the vertical (y) axis and the predictor along the horizontal (x) axis. Linear regression calculates the equation for a line that minimizes the distance between predicted scores (points on the regression line) and actual criterion scores. For example, the value of the *i*th observation would be

$$y_i = \beta_0 + \beta_1 x_i + e_i \ ,$$

where y_i is a value of the dependent variable, x_i is a value of the predictor variable, β_0 and β_1 are unknown parameters to be estimated, and e_i is an error term. As represented on the figure, β_0 would be the y-intercept and β_1 is the slope of the line. The line represents a set of predicted y_i, given a value of the predictor. Predicted y_i is often represented as

$$\hat{y}_i \ ,$$

where the "^" indicates that it is a predicted value.

The **residual**, e_i, is the difference between the observed and predicted criterion values, as shown in figure A-1:

$$e_i = y_i - \hat{y}_i \ .$$

Negative values for residuals are computed when the actual criterion score, y_i, is below the predicted y_i. Positive values are computed when the observed score exceeds the predicted. The variance of the residuals generated under model 1 is:

$$\frac{\Sigma(e_i - \bar{e})^2}{n-2} = \frac{SSE}{n-2} = MSE \ .$$

1. J. Neter and W. Wasserman. *Applied Linear Statistical Models*. Homewood, IL: Richard D. Irwin Co., 1974.

Scores with residuals large in absolute value are outliers. It is questionable whether scores with extremely large negative residuals reflect a person's true ability to perform.

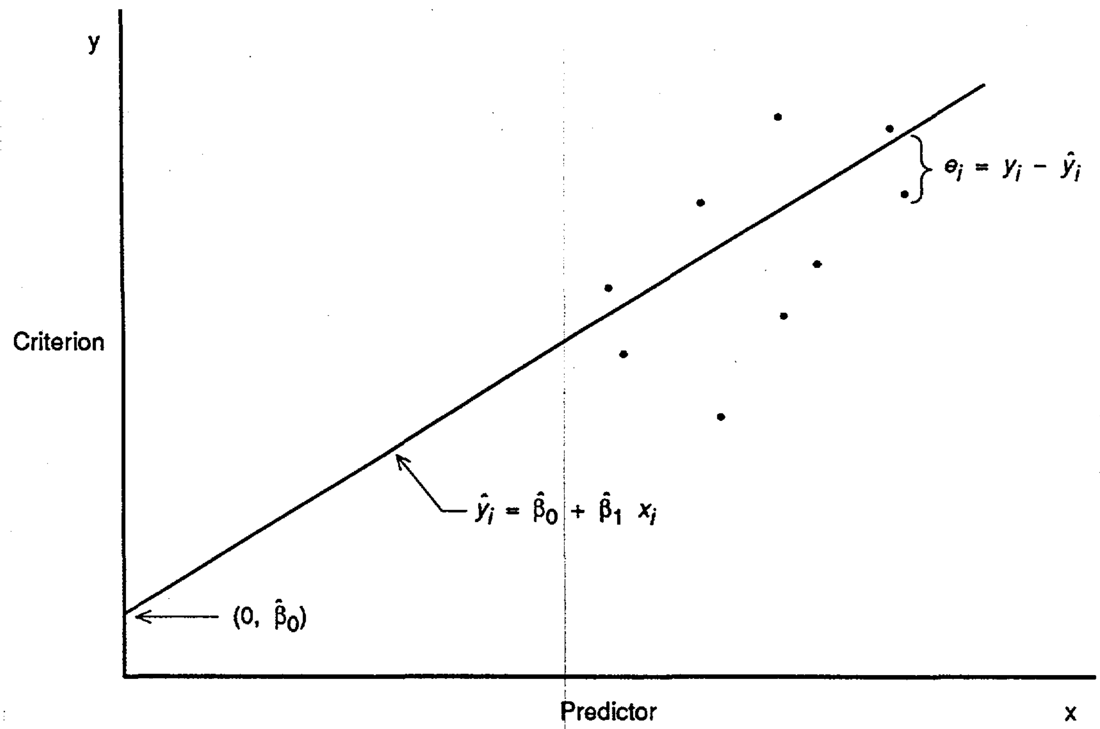


Figure A-1. Illustration of linear regression and residual analysis

To judge how extreme an outlier is, the residuals must be put on a scale that will assist in its interpretation—that is, the residual is **standardized**. Standardized residuals can be calculated by dividing e_is by the square root of their variance. The standardized residual is:

$$\frac{e_i}{(\mathrm{MSE})^{1/2}} .$$

Because this quantity has been standardized, it is expected, under the assumption of a normal distribution, that 99.74 percent of standardized residuals will fall between –3.0 and 3.0. Therefore, scores with standardized residuals below –3.0 are considered outliers.

APPENDIX B

DATA IMPUTATION PROCEDURES

APPENDIX B

DATA IMPUTATION PROCEDURES

Imputation procedures for the estimation of incomplete data are fully described in [B-1 and B-2]. Briefly, the procedure used for this study attempts to maintain the correlational structure of the original data, unlike many other imputation methods.

The initial step in the imputation procedure computes basic descriptive statistics—mean, standard deviation, minimum, maximum, and number of missing values for each variable. Intercorrelations among the variables are also computed based on all pairwise combinations of the variables; again, missing variables within each pair are noted. The variables are then ordered on the basis of their magnitude of missing data and relative intercorrelations with other variables. A stepwise regression is computed for the first variable in this ordered list that has missing data. The regression uses all prior variables in the list as predictors and stops when no further variables contribute to the prediction of the variable being imputed. The variables used for imputation included all other steps, scores for other tasks, time in service, and site.

Based on this regression, expectancy tables are constructed relating actual values to the predicted regression values. If the imputed variable is discrete, the predicted regression values are categorized into the discrete intervals of the criterion. If the imputed variable is continuous, the regressed values are categorized so that each interval contains a sufficient number of subjects. The continuous scale of the criterion is regenerated once an imputed value is determined by interpolation between the means of the regressed predicted values for adjacent categories. Table B-1 presents a hypothetical expectancy table for a discrete variable (e.g., a rating scale with values ranging from 1 to 5).

For each missing value, a predicted value is generated using the regression function, and then an "actual" value is selected randomly with probability proportional to the percentages of the expectancy table. Such a procedure yields only values that actually occurred and ensures an appropriate variation of the imputed values.

Each variable from the ordered list is processed in turn. Those variables that have imputed values are considered potential predictors in the later stepwise regressions. Once all missing values have been imputed, a second stage of imputation is conducted to determine if any variables in the later part of the ordered list would have been significant predictors of previous variables requiring imputation.

If so, the procedure is repeated for that particular variable and new imputed values are computed. In this way, any significant relationships between variables with missing values are preserved because each is used in the prediction of the other. Although it may appear that using imputed values to impute other values only builds error upon error, such redundancy is necessary to reproduce the multivariate structure of the data set. A much more complete description of the imputation procedures is provided in [B-2].

Table B-1. Expectancy table relating actual values to predicted regression values

Predicted regression value	Percentage of actual rating values at each predicted regression value				
	1	2	3	4	5
1	50	45	5		
2	15	65	20		
3		30	40	30	
4		5	20	60	15
5			20	25	55

REFERENCES

[B-1] CRM 88-259, *Analysis of Data Quality for the Infantry Phase of the Marine Corps Job Performance Measurement Project*, by Paul W. Mayberry, Mar 1989

[B-2] L. L. Wise and D. McLaughlin. *Guidebook for the Imputation of Missing Data*. Palo Alto, CA: American Institutes for Research, 1980

APPENDIX C

NUMBER OF HANDS-ON TASKS AND STEPS IMPUTED FOR EXAMINEES, MOS 6112

Table C-1. Number of hands-on tasks and steps imputed for examinees, MOS 6112 (n = 174)

	Task											
Step	0	1	2	3	4	5	6	7	8	9	Total and percentage	
0	4	0	0	0	0	0	0	0	0	0	48	(27.6)
1	0	12	0	0	0	0	0	0	0	0	12	(6.9)
2	0	4	0	0	0	0	0	0	0	0	4	(2.3)
3	0	1	16	0	0	0	0	0	0	0	17	(9.8)
4	0	2	0	16	0	0	0	0	0	0	18	(10.3)
5	0	3	0	0	2	0	0	0	0	0	5	(2.9)
6-10	0	9	3	6	6	1	0	0	0	0	25	(14.4)
11-15	0	1	1	6	5	1	0	0	0	0	14	(8.0)
16-20	0	0	4	2	8	0	0	0	0	0	14	(8.0)
21-35	0	0	0	4	5	3	0	1	0	0	13	(7.5)
36-60	0	0	0	0	2	0	1	0	0	0	3	(1.7)
>60	0	0	0	0	0	0	0	0	0	1	1	(0.6)
Total	48	32	24	34	28	5	1	1	0	1	174	(100.0)

NOTE: Entries are numbers of examinees. For example, the first entry (48) indicates that, for 48 examinees, no steps and no tasks were imputed. The next entry (12) indicates that 12 examinees had one step imputed for one task.

APPENDIX D

NUMBER OF HANDS-ON TASKS AND STEPS IMPUTED FOR EXAMINEES, MOS 6113

Table D-1. Number of hands-on tasks and steps imputed for examinees, MOS 6113 (n = 120)

	Task										Total and percentage	
Step	0	1	2	3	4	5	6	7	8	9		
0	98	0	0	0	0	0	0	0	0	0	98	(81.7)
1	0	12	0	0	0	0	0	0	0	0	12	(10.0)
2	0	1	0	0	0	0	0	0	0	0	1	(0.8)
4	0	5	0	0	0	0	0	0	0	0	5	(4.2)
5	0	0	1	0	0	0	0	0	0	0	1	(0.8)
6-10	0	0	0	1	0	0	0	0	0	0	1	(0.0)
11-15	0	2	0	0	0	0	0	0	0	0	2	(1.7)
Total	98	20	1	1	0	0	0	0	0	0	120	(100.0)

NOTE: Entries are numbers of examinees. For example, the first entry (98) indicates that, for 98 examinees, no steps and no tasks were imputed.

APPENDIX E

NUMBER OF HANDS-ON TASKS AND STEPS IMPUTED FOR EXAMINEES, MOS 6114

Table E-1. Number of hands-on tasks and steps imputed for examinees, MOS 6114 (n = 215)

	Task										Total and percentage	
Step	0	1	2	3	4	5	6	7	8	9		
0	143	0	0	0	0	0	0	0	0	0	143	(66.5)
1	0	8	0	0	0	0	0	0	0	0	8	(3.7)
2	0	3	1	0	0	0	0	0	0	0	4	(1.9)
3	0	0	0	0	0	0	0	0	0	0	0	(0.0)
4	0	1	0	0	0	0	0	0	0	0	1	(0.5)
5	0	0	0	0	0	0	0	0	0	0	0	(0.0)
6-10	0	23	2	0	0	0	0	0	0	0	25	(11.6)
11-15	0	7	5	2	0	0	0	0	0	0	14	(1.7)
16-20	0	0	1	0	0	0	0	0	0	0	1	(0.5)
21-35	0	3	6	1	0	0	0	0	0	0	10	(4.7)
36-60	0	0	0	3	4	0	0	0	0	0	7	(3.3)
>60	0	0	0	0	0	0	0	1	1	0	1	(0.5)
Total	143	45	15	6	4	0	0	1	1	0	215	(100.0)

NOTE: Entries are numbers of examinees. For example, the first entry (143) indicates that, for 143 examinees, no steps and no tasks were imputed.

APPENDIX F

NUMBER OF HANDS-ON TASKS AND STEPS IMPUTED FOR EXAMINEES, MOS 6115

Table F-1. Number of hands-on tasks and steps imputed for examinees, MOS 6115 (n = 149)

	Task										Total and	
Step	0	1	2	3	4	5	6	7	8	9	percentage	
0	115	0	0	0	0	0	0	0	0	0	115	(77.2)
1	0	19	0	0	0	0	0	0	0	0	19	(12.8)
2	0	7	0	0	0	0	0	0	0	0	7	(4.7)
3	0	4	1	0	0	0	0	0	0	0	5	(3.4)
4	0	0	0	0	0	0	0	0	0	0	0	(0.0)
5	0	0	0	0	0	0	0	0	0	0	0	(0.0)
6-10	0	1	0	0	0	0	0	0	0	0	1	(0.7)
11-15	0	0	1	0	0	0	0	0	0	0	1	(0.7)
16-20	0	0	0	0	0	0	0	0	0	0	0	(0.0)
21-35	0	0	1	0	0	0	0	0	0	0	1	(0.7)
Total	115	31	3	0	0	0	0	0	0	0	149	(100.0)

NOTE: Entries are numbers of examinees. For example, the first entry (115) indicates that, for 115 examinees, no steps and no tasks were imputed.

APPENDIX G

CORRELATIONS BEFORE AND AFTER IMPUTATION BY TASK (MOS 6112)

Table G-1. Correlations before and after imputation by task (MOS 6112, overall)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	145	.22	151	.22	145	.18	151	.17	145	.21	151	.22	145	.52	151	.51
1B	122	.06	151	.13	122	.16	151	.22	122	.31	151	.28	122	.32	151	.32
2A	150	.02	151	.02	150	.06	151	.06	150	.18	151	.18	150	.52	151	.51
2B	149	.11	151	.11	149	.11	151	.11	149	.13	151	.14	149	.62	151	.62
2C	149	.05	151	.05	149	.12	151	.12	149	.18	151	.19	149	.52	151	.53
2D	150	.11	151	.11	150	.16	151	.15	150	.44	151	.43	150	.61	151	.61
3A	147	.21	151	.19	147	.20	151	.19	147	.19	151	.20	147	.61	151	.59
3B	144	.12	151	.13	144	.06	151	.04	144	.08	151	.04	144	.46	151	.44
3C	149	.11	151	.12	149	.06	151	.06	149	.11	151	.12	149	.57	151	.57
4A	147	.14	151	.14	147	.07	151	.06	147	.21	151	.22	147	.51	151	.50
4B	129	-.01	151	.00	129	.12	151	.12	129	.23	151	.27	129	.56	151	.54
5A	134	.04	151	.09	134	.04	151	.06	134	.31	151	.31	134	.47	151	.45
5B	148	.15	151	.16	148	.27	151	.28	148	.40	151	.40	148	.50	151	.50
6A[a]	93	.15	151	.24	93	.15	151	.23	93	.26	151	.29	93	.37	151	.45
7A	149	.10	151	.10	149	.22	151	.22	149	.33	151	.33	149	.62	151	.62
7B	147	.21	151	.23	147	.28	151	.31	147	.18	151	.20	147	.41	151	.42
8A	149	.13	151	.15	149	.11	151	.13	149	.25	151	.26	149	.39	151	.37
8B[a]	97	.19	151	.15	97	.26	151	.17	97	.21	151	.16	97	.39	151	.13
8C[a]	75	.12	151	.06	75	.31	151	.18	75	.35	151	.31	75	.53	151	.39

a. Large changes in correlation for these tasks are the result of very few complete data at site B. For task 6A, only 15 site B cases were complete. For task 8C, only two cases were complete at site B. For task 8B, only 16 cases were complete at site B. These numbers are for listwise deletion.

Table G-2. Correlations before and after imputation by task (MOS 6112, site A)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	82	.22	85	.22	82	.16	85	.17	82	.22	85	.24	82	.37	85	.39
1B	72	-.05	85	.00	72	.01	85	.11	72	.46	85	.44	72	.39	85	.37
2A	85	.13	85	.13	85	.22	85	.22	85	.27	85	.27	85	.32	85	.32
2B	83	.04	85	.06	83	.21	85	.21	83	.23	85	.24	83	.39	85	.40
2C	83	.12	85	.13	83	.21	85	.24	83	.28	85	.29	83	.35	85	.36
2D	84	.01	85	.00	84	.07	85	.06	84	.44	85	.43	84	.63	85	.62
3A	82	.18	85	.16	82	.26	85	.27	82	.30	85	.32	82	.36	85	.37
3B	84	.04	85	.05	84	.14	85	.14	84	.27	85	.27	84	.42	85	.43
3C	85	.12	85	.12	85	.15	85	.15	85	.22	85	.22	85	.31	85	.31
4A	83	.21	85	.21	83	.15	85	.15	83	.29	85	.29	83	.39	85	.39
4B	84	.00	85	.00	84	.21	85	.22	84	.27	85	.28	84	.48	85	.49
5A	79	.04	85	.08	79	.01	85	.05	79	.26	85	.30	79	.49	85	.49
5B	83	.25	85	.25	83	.38	85	.38	83	.37	85	.37	83	.56	85	.55
6A	78	.11	85	.11	78	.15	85	.16	78	.31	85	.31	78	.44	85	.42
7A	83	.14	85	.14	83	.26	85	.27	83	.37	85	.37	83	.55	85	.55
7B	82	.15	85	.20	82	.22	85	.28	82	.24	85	.28	82	.39	85	.47
8A	83	.29	85	.31	83	.20	85	.24	83	.23	85	.26	83	.35	85	.35
8B	81	.10	85	.11	81	.20	85	.20	81	.27	85	.27	81	.44	85	.40
8C	73	.15	85	.15	73	.31	85	.29	73	.33	85	.38	73	.61	85	.62

Table G-3. Correlations before and after imputation by task (MOS 6112, site B)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	63	.33	66	.32	63	.29	66	.27	63	.23	66	.22	63	.54	66	.51
1B	50	.05	66	.14	50	.16	66	.24	50	.20	66	.19	50	.55	66	.54
2A	65	.13	66	.12	65	.15	66	.12	65	.18	66	.16	65	.49	66	.46
2B	66	.41	66	.41	66	.32	66	.32	66	.11	66	.11	66	.61	66	.61
2C	66	.22	66	.22	66	.28	66	.28	66	.15	66	.15	66	.43	66	.43
2D	66	.23	66	.23	66	.28	66	.28	66	.44	66	.44	66	.65	66	.65
3A	65	.45	66	.45	65	.39	66	.39	65	.16	66	.16	65	.59	66	.58
3B	60	.29	66	.28	60	.16	66	.12	60	.03	66	-.05	60	.39	66	.36
3C	64	.31	66	.30	64	.20	66	.18	64	.08	66	.08	64	.58	66	.57
4A	64	.28	66	.29	64	.17	66	.17	64	.19	66	.21	64	.41	66	.39
4B	45	.18	66	.12	45	.22	66	.18	45	.27	66	.28	45	.60	66	.45
5A	55	.11	66	.16	55	.15	66	.14	55	.37	66	.34	55	.40	66	.35
5B	65	.08	66	.09	65	.17	66	.18	65	.44	66	.45	65	.53	66	.54
6A[a]	15	.40	66	.36	15	.14	66	.30	15	-.09	66	.27	15	.34	66	.58
7A	66	.23	66	.23	66	.36	66	.36	66	.33	66	.33	66	.55	66	.55
7B	65	.33	66	.32	65	.42	66	.40	65	.10	66	.08	65	.38	66	.39
8A	66	.08	66	.08	66	.08	66	.08	66	.27	66	.27	66	.35	66	.35
8B[a]	16	.43	66	.09	16	.45	66	.04	16	.01	66	.05	16	.85	66	.21
8C[a]	2	–	66	-.03	2	–	66	.07	2	–	66	.23	2	–	66	.32

a. Too few complete cases existed to make the original correlation meaningful.

APPENDIX H

CORRELATIONS BEFORE AND AFTER IMPUTATION BY TASK (MOS 6113)

Table H-1. Correlations before and after imputation by task (MOS 6113, overall)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*
1A	88	.07	88	.07	88	.24	88	.24	88	.29	88	.29	88	.51	88	.51
1B	87	.23	88	.23	87	.22	88	.22	87	.24	88	.25	87	.45	88	.46
1C	88	.08	88	.08	88	.24	88	.24	88	.32	88	.32	88	.47	88	.47
1D	88	.09	88	.09	88	.23	88	.23	88	.46	88	.46	88	.65	88	.65
2A	87	.14	88	.14	87	.33	88	.32	87	.36	88	.35	87	.58	88	.57
3A	87	.17	88	.15	87	.18	88	.17	87	.42	88	.43	87	.62	88	.62
3B	87	.08	88	.08	87	.08	88	.08	87	.30	88	.30	87	.51	88	.51
4A	88	–.05	88	-.05	88	.16	88	.16	88	.11	88	.11	88	.41	88	.41
4B	88	.12	88	.12	88	.42	88	.42	88	.39	88	.39	88	.66	88	.66
5A	88	.21	88	.21	88	.25	88	.25	88	.15	88	.15	88	.32	88	.32
5B	87	.17	88	.17	87	.37	88	.37	87	.20	88	.20	87	.42	88	.42
6A	79	.30	88	.26	79	.36	88	.41	79	.22	88	.24	79	.60	88	.62
6B	88	.21	88	.21	88	.27	88	.27	88	.40	88	.40	88	.59	88	.59
7A	83	.19	88	.21	83	.37	88	.33	83	.37	88	.36	83	.59	88	.58
7B	87	.28	88	.28	87	.29	88	.29	87	.31	88	.31	87	.29	88	.30
8A	88	.16	88	.16	88	.20	88	.20	88	.20	88	.20	88	.35	88	.35

NOTE: For MOS 6113, testing occurred only at site B.

APPENDIX I

CORRELATIONS BEFORE AND AFTER IMPUTATION BY TASK (MOS 6114)

Table I-1. Correlations before and after imputation by task (MOS 6114, overall)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	176	.22	179	.22	176	.38	179	.37	176	.21	179	.21	176	.47	179	.47
1B	171	.11	179	.13	171	.29	179	.32	171	.15	179	.16	171	.52	179	.55
2A	175	.14	179	.14	175	.24	179	.24	175	.14	179	.14	175	.39	179	.39
3A	177	.19	179	.19	177	.21	179	.21	177	.17	179	.17	177	.58	179	.58
3B	176	.16	179	.16	176	.22	179	.21	176	.18	179	.17	176	.50	179	.49
4A	175	.27	179	.27	175	.35	179	.36	175	.19	179	.17	175	.64	179	.64
5A	173	.16	179	.14	173	.25	179	.24	173	.08	179	.07	173	.39	179	.39
5B	164	.13	179	.15	164	.20	179	.22	164	.09	179	.09	164	.52	179	.52
6A	176	.21	179	.20	176	.26	179	.23	176	.10	179	.07	176	.46	179	.45
6B	177	.14	179	.13	177	.17	179	.17	177	.09	179	.09	177	.31	179	.30
7A	172	.19	179	.18	172	.32	179	.31	172	.27	179	.27	172	.58	179	.57
7B	173	.22	179	.20	173	.19	179	.17	173	.10	179	.10	173	.44	179	.42
8A	179	.13	179	.13	179	.21	179	.21	179	.09	179	.09	179	.41	179	.41
8B	150	.18	179	.13	150	.15	179	.14	150	.25	179	.28	150	.51	179	.50
8C	166	.13	179	.14	166	.18	179	.20	166	.07	179	.09	166	.41	179	.42

Table I-2. Correlations before and after imputation by task (MOS 6114, site A)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	48	.02	49	.02	48	.27	49	.24	48	.28	49	.25	48	.41	49	.37
1B	49	.10	49	.10	49	.28	49	.28	49	.08	49	.08	49	.53	49	.53
2A	46	.25	49	.25	46	.27	49	.31	46	.00	49	.03	46	.55	49	.54
3A	49	.02	49	.02	49	-.01	49	-.01	49	.12	49	.12	49	.44	49	.44
3B	48	.11	49	.11	48	.06	49	.06	48	-.10	49	-.10	48	.23	49	.23
4A	47	.04	49	.07	47	.07	49	.09	47	-.14	49	-.16	47	.53	49	.51
5A	47	.23	49	.19	47	.32	49	.27	47	-.11	49	-.12	47	.28	49	.26
5B	44	.08	49	.09	44	.15	49	.17	44	-.02	49	.00	44	.13	49	.16
6A	48	.06	49	.06	48	.10	49	.10	48	-.27	49	-.26	48	.38	49	.39
6B	48	.11	49	.07	48	.05	49	.06	48	.02	49	.04	48	.28	49	.27
7A	47	.23	49	.24	47	.29	49	.29	47	.13	49	.13	47	.76	49	.76
7B	48	.12	49	.14	48	.21	49	.16	48	-.10	49	-.03	48	.38	49	.38
8A	49	.06	49	.06	49	.06	49	.06	49	-.05	49	-.05	49	.28	49	.28
8B	46	.03	49	-.03	46	.10	49	.09	46	.13	49	.25	46	.50	49	.57
8C	48	.13	49	.13	48	.21	49	.21	48	.09	49	.07	48	.52	49	.51

Table I-3. Correlations before and after imputation by task (MOS 6114, site B)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	128	.27	130	.27	128	.39	130	.39	128	.19	130	.19	128	.55	130	.55
1B	122	.06	130	.09	122	.25	130	.28	122	.13	130	.15	122	.39	130	.43
2A	129	.10	130	.10	129	.22	130	.22	129	.19	130	.19	125	.43	130	.43
3A	128	.16	130	.16	128	.13	130	.14	128	.12	130	.13	128	.36	130	.36
3B	128	.13	130	.13	128	.17	130	.16	128	.20	130	.19	128	.37	130	.36
4A	128	.30	130	.29	128	.35	130	.36	128	.24	130	.24	128	.60	130	.60
5A	126	.12	130	.10	126	.21	130	.19	126	.06	130	.05	126	.29	130	.28
5B	120	.09	130	.08	120	.13	130	.12	120	-.01	120	.00	120	.31	130	.30
6A	128	.25	130	.22	128	.26	130	.23	128	.26	130	.20	128	.45	130	.42
6B	129	.13	130	.13	129	.17	130	.16	129	.09	130	.09	129	.25	130	.25
7A	125	.15	130	.13	125	.30	130	.28	125	.32	130	.32	125	.51	130	.49
7B	125	.22	130	.20	125	.14	130	.13	125	.12	130	.11	125	.39	130	.38
8A	130	.11	130	.11	130	.19	130	.19	130	.11	130	.11	130	.33	130	.33
8B	104	.18	130	.14	104	.10	130	.09	104	.26	130	.28	104	.43	130	.42
8C	118	.11	130	.12	118	.15	130	.17	118	.03	130	.07	118	.38	130	.38

APPENDIX J

CORRELATIONS BEFORE AND AFTER IMPUTATION BY TASK (MOS 6115)

Table J-1. Correlations before and after imputation by task (MOS 6115, overall)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	114	.08	114	.08	114	.04	114	.04	114	.24	114	.24	114	.61	114	.61
1B	113	-.01	114	.00	113	.04	114	.03	113	.23	114	.24	113	.62	114	.62
1C	114	-.03	114	-.03	114	.10	114	.10	114	.21	114	.21	114	.47	114	.47
1D	114	.07	114	.07	114	.22	114	.22	114	.27	114	.27	114	.33	114	.33
2A	112	.23	114	.22	112	.14	114	.13	112	.18	114	.17	112	.56	114	.56
3A	113	.16	114	.17	113	.13	114	.13	113	.15	114	.16	113	.68	114	.68
3B	112	.11	114	.12	112	.13	114	.14	112	.12	114	.12	112	.68	114	.67
4A	111	.07	114	.09	111	-.01	114	.00	111	.26	114	.26	111	.63	114	.63
4B	114	.09	114	.09	114	.11	114	.11	114	.24	114	.24	114	.49	114	.49
5A	110	.25	114	.24	110	.18	114	.18	110	.14	114	.14	110	.60	114	.60
6A	113	.13	114	.12	113	.00	114	.00	113	-.05	114	-.05	113	.50	114	.49
6B	104	.12	114	.12	104	.09	114	.10	104	.15	114	.16	104	.72	114	.71
7A	113	.03	114	.03	113	.03	114	.03	113	.02	114	.02	113	.55	114	.55
7B	114	.11	114	.11	114	.10	114	.10	114	.14	114	.14	114	.72	114	.72
7C	113	-.01	114	.00	113	.05	114	.06	113	.19	114	.19	113	.61	114	.61
8A	110	.09	114	.10	110	.00	114	.00	110	.12	114	.12	110	.53	114	.52

Table J-2. Correlations before and after imputation by task (MOS 6115, site A)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	48	.15	48	.15	48	.14	48	.14	48	.28	48	.28	48	.51	48	.51
1B	48	.09	48	.09	48	.14	48	.14	48	.31	48	.31	48	.66	48	.66
1C	48	.12	48	.12	48	.30	48	.30	48	.26	48	.26	48	.52	48	.52
1D	48	.03	48	.03	48	.23	48	.23	48	.26	48	.26	48	.43	48	.43
2A	47	.27	48	.27	47	.27	48	.27	47	.20	48	.19	47	.24	48	.24
3A	48	.12	48	.12	48	.08	48	.08	48	.29	48	.29	48	.34	48	.34
3B	47	-.07	48	-.01	47	.12	48	.23	47	.02	48	.05	47	.21	48	.24
4A	46	-.11	48	-.08	46	-.03	48	-.02	46	.24	48	.24	46	.47	48	.47
4B	48	.05	48	.05	48	.07	48	.07	48	.36	48	.36	48	.51	48	.51
5A	48	.41	48	.41	48	.31	48	.31	48	.08	48	.08	48	.23	48	.23
6A	48	.08	48	.08	48	-.10	48	-.10	48	-.03	48	-.03	48	-.02	48	-.02
6B	48	.08	48	.08	48	.13	48	.13	48	.39	48	.39	48	.34	48	.34
7A	48	.16	48	.16	48	.05	48	.05	48	.25	48	.25	48	.42	48	.42
7B	48	.21	48	.21	48	.24	48	.24	48	.13	48	.13	48	.18	48	.18
7C	48	.08	48	.08	48	.25	48	.25	48	.43	48	.43	48	.54	48	.54
8A	45	.00	48	.02	45	.09	48	.09	45	.09	48	.11	45	.25	48	.27

Table J-3. Correlations before and after imputation by task (MOS 6115, site B)

	ENL-MM				CAT-MM				TIS				PCTOT			
	Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation		Before imputation		After imputation	
Task	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
1A	66	.05	66	.05	66	.02	66	.02	66	.21	66	.21	66	.35	66	.35
1B	65	-.07	66	-.06	65	.00	66	-.02	65	.17	66	.18	65	.49	66	.48
1C	66	-.10	66	-.10	66	.04	66	.04	66	.18	66	.18	66	.30	66	.30
1D	66	.11	66	.11	66	.20	66	.20	66	.29	66	.29	66	.58	66	.58
2A	65	.26	66	.25	65	.15	66	.15	65	.17	66	.17	65	.45	66	.45
3A	65	.21	66	.22	65	.21	66	.22	65	.13	66	.13	65	.66	66	.66
3B	65	.17	66	.17	65	.17	66	.17	65	.12	66	.12	65	.64	66	.64
4A	65	.17	66	.16	65	.03	66	.02	65	.29	66	.28	65	.45	66	.44
4B	66	.11	66	.11	66	.16	66	.16	66	.18	66	.18	66	.33	66	.33
5A	62	.22	66	.21	62	.17	66	.17	62	.16	66	.15	62	.64	66	.64
6A	65	.18	66	.16	65	.08	66	.06	65	-.17	66	-.18	65	.20	66	.20
6B	56	.12	66	.16	56	.15	66	.17	56	.09	66	.09	56	.58	66	.59
7A	65	-.03	66	-.03	65	.05	66	.04	65	-.15	66	-.15	65	.40	66	.40
7B	66	.10	66	.10	66	.14	66	.14	66	.12	66	.12	66	.55	66	.55
7C	65	-.10	66	-.07	65	-.08	66	-.07	65	-.01	66	-.02	65	.45	66	.45
8A	65	.13	66	.13	65	-.01	66	-.01	65	.09	66	.10	65	.44	66	.43